ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

IDENTIFICATION AND LOCALIZATION OF HIGH IMPEDANCE FAULTS IN DISTRIBUTION NETWORKS

Ph.D. THESIS

Eren BAHARÖZÜ

Department of Electrical Engineering

Electrical Engineering Programme

MARCH 2025

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

IDENTIFICATION AND LOCALIZATION OF HIGH IMPEDANCE FAULTS IN DISTRIBUTION NETWORKS

Ph.D. THESIS

Eren BAHARÖZÜ
(504182003)

Department of Electrical Engineering

Electrical Engineering Programme

Thesis Advisor: Assoc. Prof. Dr. Suat İLHAN
Thesis Co-Advisor: Asst. Prof. Dr. Gürkan SOYKAN

MARCH 2025

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

DAĞITIM ŞEBEKELERİNDEKİ YÜKSEK EMPEDANS ARIZALARIN TESPİTİ VE KONUMUNUN BELİRLENMESİ

DOKTORA TEZİ

Eren BAHARÖZÜ
(504182003)

Elektrik Mühendisliği Anabilim Dalı

Elektrik Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Suat İLHAN
Eş Danışman: Dr. Öğr. Üyesi Gürkan SOYKAN

MART 2025

Eren BAHARÖZÜ, a Ph.D. student of ITU Graduate School student ID 504182003, successfully defended the thesis/dissertation entitled "IDENTIFICATION AND LOCALIZATION OF HIGH IMPEDANCE FAULTS IN DISTRIBUTION NETWORKS", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Assoc. Prof. Dr. Suat İLHAN** Istanbul Technical University	.....
Co-advisor :	**Asst. Prof. Dr. Gürkan SOYKAN** Bahcesehir University	.....
Jury Members :	**Prof. Dr. Mustafa BAĞRIYANIK** Istanbul Technical University	.....
	Prof. Dr. Özcan KALENDERLI Istanbul Technical University	.....
	Assoc. Prof. Dr. Oktay ARIKAN Yıldız Technical University	.....
	Prof. Dr. Şükran Emel ÖNAL Istanbul Technical University	.....
	Prof. Dr. Mukden UĞUR Turkish German University	.....

Date of Submission : 10 January 2025
Date of Defense : 17 March 2025

To my beloved sister Sıla Pehlivanoğlu,

FOREWORD

I would like to start by expressing my gratitude to Assoc. Prof. Suat ILHAN and Dr. Gürkan SOYKAN for their advice, support, and guidance throughout the entire period of my thesis. I also would like to thank my committee members for their valuable suggestions and comments. Additionally, I would like to honor Prof. Dr. Okan ÖZGÖNENEL, who served on the thesis monitoring jury and whom we sadly lost on October 29, 2024.

Lastly, I would like to thank my family and my beloved wife, who have never withheld their support and encouragement from me during my thesis work.

March 2025 Eren BAHARÖZÜ

TABLE OF CONTENTS

Page

ABBREVIATIONS

A : Approximation
ADDT : Advanced Distortion Detection Technique
ANFIS : Adaptive Neuro-Fuzzy Inference System
ANN : Artificial Neural Network
BP : Backpropagation
cm : Centimeter
CNN : Convolutional Neural Network
CWT : Continuous Wavelet Transform
D : Detail
DC : Direct Current
DT : Decision Tree
DWT : Discrete Wavelet Transform
EEMD : Ensemble Empirical Mode Decomposition
FFT : Fast Fourier Transform
FT : Fourier Transform
GA : Genetic Algorithm
GPR : Gaussian Process Regression
HHT : Hilbert–Huang Transform
HIF : High Impedance Fault
HV : High Voltage
Hz : Hertz
IIPV : Integrated Inner Product Value
KNN : K-Nearest Neighbors
MHz : Megahertz
MODWT : Maximal Overlapping Discrete Wavelet
m : Meter
ms : Millisecond
NM : Not Modeled
NN : Neural Network

NS : Not Stated

PSO : Particle Swarm Optimization

ReLU : Rectified Linear Unit

RBF : Radial Basis Functions

RMS : Root Mean Square

s : Second

SNR : Signal to Noise Ratio

SOMN : Self Organizing Mapping Neural Network

STFT : Short Time Fourier Transform

SVM : Support Vector Machine

TACS : Transient Analysis of Control Systems

TKEO : Teager-Kaiser Energy Operator

WLS : Weighted Least Squares

WNN : Wavelet Neural Network

WPT : Wavelet Packet Transform

WT : Wavelet Transform

XGBoost : Extreme Gradient Boosting

ZSC : Zero Sequence Current

SYMBOLS

a	**:** Scaling parameter
b	**:** Translation parameter
C	**:** Capacitance
ω	**:** Weight between input and hidden layer of ANN
α	**:** Weight between hidden and output layer of ANN
t	**:** Time
Ω	**:** Ohm
$\mathbf{X_i}$	**:** i^{th} input of the data
Y	**:** Output of ANN
φ(n)	**:** Mother Wavelet
L	**:** Inductance
R1	**:** First Resistance
R2	**:** Second Resistance
σ	**:** Standard Deviation
$\mathbf{\sigma^2}$	**:** Variance
μ	**:** Mean
N	**:** The Number of The Data Points
Sk	**:** Skewness
Kr	**:** Kurtosis
Ex	**:** Sum of Energy Content

LIST OF TABLES

Page

LIST OF FIGURES

Page

IDENTIFICATION AND LOCALIZATION OF HIGH IMPEDANCE FAULTS IN DISTRIBUTION NETWORKS

SUMMARY

High impedance faults continue to pose significant challenges in ensuring reliable power systems due to their low fault currents and complex characteristics. HIFs can occur in various scenarios, such as when a live conductor breaks and either contacts the ground or nearby trees touching an energized overhead line conductor. These faults usually generate fault currents with lower amplitudes than the threshold values set for protection systems and are largely undetectable by traditional protection equipment. Due to their arcing nature and difficulty in detection and location, high impedance faults have the potential to cause fires and damage network equipment. Therefore, they pose serious risks not only for energy distribution companies but also for human and environmental safety.

This situation has prompted researchers to focus on high impedance faults for over forty years. To date, numerous studies have been conducted to explore and mitigate HIF-related problems caused by high impedance faults. However, a universal solution for detecting or locating HIFs remains absent. Therefore, an economical and applicable solution, especially for the localization of HIFs, was sought within the thesis while enhancing the understanding of HIFs.

In the early years, when the subject was studied, there was limited information available about the behaviors of HIFs. This led all early studies on the subject to use signals obtained from laboratory tests or staged faults in the field. Through these efforts, researchers were able to analyze actual fault signals and identify the key characteristics of HIFs. Once these unique behaviors were recognized, models that replicate one or more behaviors of HIFs were proposed. Most subsequent studies have utilized these proposed models to identify or locate HIFs. There are two main reasons for this. First, conducting field and laboratory tests can be quite challenging. The other is that many scenarios can be explored within a simulation environment using HIF models, which can be easily adapted to fit various distribution networks through simple modifications. However, it is essential to employ accurate and complete models to examine clearly the influences of such faults. Thus, the thesis presents a detailed explanation of HIF characteristics and proposed HIF models with their represented behaviors in chronological order after the introduction section.

After examining and clearly understanding the fundamental concepts of HIF, the thesis shifts its focus to studies related to the detection and localization of HIFs within power systems. Comprehensive literature reviews were conducted for both topics. The proposed methods were described in detail, considering all key indices, such as system changes, noise conditions, test networks, and the HIF model. In addition, detection and localization techniques were categorized and compared within their classes, emphasizing their pros and cons. Finally, outcomes from the analyzed articles were addressed under the literature review section.

In the third section of the thesis, the HIF experiments carried out in Istanbul Technical University Fuat Külünk High Voltage Laboratory are described in detail. While the main characteristics of HIF have been outlined in existing literature, it is crucial to assess whether these fault characteristics are observed consistently in real fault scenarios. Therefore, extensive laboratory experiments on HIF were designed and executed on various contact surfaces to capture realistic HIF signals. Due to the limitations of the experimental setup, the amplitudes of HIF currents obtained in the laboratory are typically lower than those of a fault observed in the field. However, by scaling the relevant signals as described in Section 4 and using them in simulations, the drawbacks of utilizing signals obtained from experimental tests have been eliminated. Through the analysis of the fault data collected from the laboratory measurements, a deeper understanding of the nature of HIF was achieved while the documented characteristics of HIF were validated.

After performing a literature review and laboratory experiments, two different feature approaches for HIF detection were tested and compared by utilizing HIF signals captured in the laboratory. According to the literature, most published detection methods have yielded promising outcomes, while machine learning methods have demonstrated superior accuracy. Therefore, in addition to employing a different set of input features, two distinct machine-learning algorithms were analyzed in the thesis. Furthermore, events exhibiting similar characteristics to HIFs, along with HIF signals obtained in the laboratory and modeled, were included in the simulations to verify the effectiveness of the methods. The study also examined how different configurations of machine learning techniques, various mother waves, and varying noise levels affect the results. Findings indicated that the ANN algorithm produced more accurate detection results for noisy and clean signals, while the standard deviation approach used as input features outperformed the energy sum approach. However, it was observed that both using HIF signals obtained in a laboratory environment and adding noise to the signals caused a reduction in the overall performance of the methods.

After studying distinguishing HIFs from other faults, which is the first step before localization, the thesis focuses on pinpointing the location of HIFs accurately. According to the conducted studies, there have been relatively few studies on locating HIFs compared to distinguishing them. Approximately 80% of studies have employed either impedance-based methods or machine learning techniques to pinpoint the exact location of HIFs. In recent years, machine learning methods that utilize features extracted from voltage and current signals have gained popularity due to their lower error rates. However, none of these studies have implemented any feature selection method or provided a rationale for their chosen features. Therefore, it is aimed to investigate the influence of feature selection techniques on HIF localization in the thesis. In this context, firstly, current waveforms of high impedance faults were obtained with the help of simulation studies. Then, a seven-level discrete wavelet transform was applied to the current signals, and five statistical parameters were calculated for each wavelet coefficient, resulting in a total of 40 features. These statistical parameters are standard deviation, variance, skewness, kurtosis, and the sum of energy content. Feature selection methods were then applied to the created feature pool, and the performance of ANNs trained with the selected feature subsets was evaluated. The features proposed from each feature selection method were tested 250 times on artificial neural networks, and the averages of all results were taken into account in the evaluations. Based on the comparisons, wrapper methods outperformed others when the ANN settings were the same. However, features chosen by the random

forest method consistently achieved lower error distances across different ANN configurations and demonstrated higher accuracies in test data, even under noisy conditions. Overall, all feature selection methods successfully reduced computational complexity and enhanced accuracy, even in noisy conditions. Furthermore, it was observed that these methods could successfully identify significant and irrelevant features within the data.

Following an assessment of available detection and localization methods by aiming to complete the missing areas in existing publications, this thesis proposes a new data-driven HIF location method that can identify the faulty branch associated with HIF and accurately pinpoint its location in a distribution network. This innovative method utilizes the traveling wave approach, Discrete Wavelet Transform (DWT), and machine learning technique. Since ANN is used commonly in localization studies, the proposed methodology incorporates this method as the chosen machine learning method. Normally, estimating the fault distance accurately using traveling wave methods requires identifying each reflected and refracted traveling wave. However, using this method directly can be challenging, especially in distribution networks with multiple short branches. Therefore, the proposed method utilizes information from a certain number of selected traveling waves that can be easily distinguished from the current signal in ANN. By employing this selection process alongside a combination of traveling wave and machine learning techniques, the aim was to overcome certain weaknesses of each method. The proposed approach has been evaluated across multiple scenarios, considering different inception angles, machine learning techniques, load variations, extreme conditions, and distribution networks to validate its robustness and efficiency. The findings indicate that the proposed method exhibits high accuracy in detecting the faulty network branch and has a minimal error rate in calculating the distance of the fault location.

DAĞITIM ŞEBEKELERİNDEKİ YÜKSEK EMPEDANS ARIZALARIN TESPİTİ VE KONUMUNUN BELİRLENMESİ

ÖZET

Yüksek empedanslı arızalar, düşük arıza akımları ve karmaşık özellikleri nedeniyle güvenilir güç sistemlerinin sağlanmasında önemli zorluklar oluşturmaya devam etmektedir. Yüksek empedans arızası, enerjili bir havai hat iletkeninin kopup zemine temas etmesi veya yakındaki ağaçların enerjili bir havai hat iletkenine temas etmesi gibi çeşitli senaryolarda meydana gelebilmektedir. Bu arızalar, genellikle koruma sistemleri için belirlenen eşik değerlerinden daha düşük genlikli arıza akımları üretmekte olup geleneksel koruma donanımlarına tarafından büyük oranda tespit edilememektedirler. Yüksek empedans arızalarının, ark ile ilişkili olması ve tespiti ile konumunun zor belirlenebilmesi nedeniyle yangın çıkarma ve şebeke donanımlarına zarar verme potansiyeline sahiptirler. Bu nedenle sadece enerji dağıtım şirketleri için değil insan ve çevre güvenliği açısından ciddi riskler oluşturmaktadır.

Bu durum araştırmacıları kırk yıldan uzun süredir yüksek empedans arızalara odaklanmaya yöneltmiştir. Günümüze kadar yüksek empedans arızası kaynaklı sorunları tespit edip, azaltmayı amaçlayan çok sayıda çalışma yapılmıştır. Ancak, bu arızaları tespit etmek veya konumunu belirlemek için halen evrensel bir çözüm bulunamamıştır. Bu nedenle, tez kapsamında bir yandan yüksek empedans arızalarının daha iyi anlaşılması sağlarken bir yandan da özellikle konumunun noktasal olarak belirlenmesi için ekonomik ve uygulanabilir bir çözüm aranmıştır.

Konunun çalışıldığı ilk yıllarda yüksek empedans arızalarının davranışları hakkında bilgi eksiklikleri bulunmaktaydı. Bu da konuyla ilgili yapılan tüm erken çalışmaların laboratuvar testlerinden veya sahada bilerek oluşturulan arızalardan elde edilen sinyallerin kullanılması yol açtı. Bu çalışmalar sayesinde araştırmacılar gerçek arıza sinyallerini analiz edebildiler ve yüksek empedans arızalarının temel özelliklerini belirleyebildiler. Bu benzersiz davranışlar tespit edildikten sonra ise, yüksek empedans arızalarının özelliklerinden bir veya birkaçını yansıtacak modellemeler önerildi. Sonraki ve günümüzde yapılan çalışmaların birçoğunda yüksek empedans arızalarını tespit edebilmek veya yerini belirlemek için önerilen bu modeller kullanıldı. Bunun iki temel nedeni bulunmaktadır. İlki saha ve laboratuvar testlerinin yapılmasındaki zorluklardır. Diğeri ise birçok durumun yüksek empedans arıza modelleri kullanılarak simülasyon ortamında incelenebilmesi ve yapılacak basit modifikasyonlar sayesinde yüksek empedans arıza modellerinin her dağıtım şebekesine uyum sağlayacak şekilde uyarlanabilmesidir. Ancak bu tür arızalarının etkilerini net bir şekilde incelemek için doğru ve tamamlanmış modellerin kullanılması önemlidir. Bu nedenle tezde, öncelikle yüksek empedans arızalarının özelliklerinin ayrıntılı bir açıklaması yapılmış ve önerilen yüksek empedans arıza modelleri yansıttığı özellikler ile beraber kronolojik olarak sunulmuştur.

Yüksek empedans arızaları ile ilgili temel kavramlar incelenip net bir şekilde anlaşıldıktan sonra, tezde güç sistemlerinde yüksek empedans arızalarının tespiti ve konumunun belirlenmesiyle alakalı çalışmalara odaklanılmıştır. Her iki konu için de literatür kapsamlı bir şekilde araştırılmıştır. Önerilen yöntemlerde değerlendirilen sistem değişiklikleri, gürültü durumu, test şebekeleri ve kullanılan yüksek empedans arıza modeli gibi tüm önemli endeksler dikkate alınarak ayrıntılı olarak incelenmiştir. Ayrıca, tespit ve konum belirleme teknikleri artıları ve eksileri vurgulanarak kendi sınıfları içinde kategorize edilmiş ve karşılaştırılır. Son olaraksa, literatür incelemesi başlığı altında incelenen makalelerden elde edilen sonuçlar değerlendirilmiştir.

Tezin üçüncü bölümünde ise, İstanbul Teknik Üniversitesi Fuat Külünk Yüksek Gerilim Laboratuvarı'nda gerçekleştirilen yüksek empedans arızası deneyleri ayrıntılı olarak anlatılmıştır. Yüksek empedans arızalarının temel özellikleri mevcut literatürde ana hatlarıyla belirtilmiş olsa da, bu arıza özelliklerinin gerçeğe yakın arıza senaryolarında da tutarlı bir şekilde gözlenip gözlenmediğini değerlendirmek oldukça önemlidir. Bu nedenle, gerçeğe yakın yüksek empedans arızalarının sinyallerini yakalamak için çeşitli temas yüzeylerinde kapsamlı laboratuvar deneyleri tasarlanmış ve yürütülmüştür. Deneysel kurulumun sınırlamaları nedeniyle laboratuvarda elde edilen yüksek empedans arıza akımlarının genlikleri genellikle sahada gözlemlenen bir arızanınkinden daha düşük olmaktadır. Fakat ilgili sinyallerin 4. Başlıkta açıklandığı gibi ölçekleme yapılarak simülasyonlarda kullanılmaları sayesinde, deneysel testlerden elde edilen sinyallerin kullanılmasındaki bu dezavantaj ortadan kaldırmıştır. Laboratuvar ölçümlerinden toplanan arıza verilerinin analizi ile de yüksek empedans arızasının doğası hakkında daha derin bir anlayış elde edilirken, bu tip arızalarının belgelenen özellikleri de doğrulanmıştır.

Literatür taraması ve laboratuvar deneyleri yapıldıktan sonra, laboratuvarda kayıt altına alınan yüksek empedans arıza sinyallerinden de yararlanılarak yüksek empedans arızaların tespitinde kullanılmasını önerilen iki farklı özellik yaklaşımı test edildi ve karşılaştırıldı. Literatüre göre, yayınlanan tespit yöntemlerinin çoğu ümit verici sonuçlar verirken, makine öğrenimi yöntemleri üstün doğruluk göstermiştir. Bu nedenle, değişik özellik kümelerinin kullanılmasına ek olarak iki farklı makine öğrenimi algoritmasını da tez içerisinde inceledi. Ayrıca, yüksek empedans arızalarına benzer özellikler gösteren olaylar, laboratuvar ortaında elde edilmiş ve modellenmiş yüksek empedans arıza sinyalleriyle birlikte, yöntemlerin etkinliğini doğrulamak için simülasyonlara dahil edildi. Çalışmada ayrıca, makine öğrenmesi tekniklerinin farklı yapılandırmalarının, çeşitli ana dalgacıkların ve değişen gürültü seviyelerinin sonuçları nasıl etkilediği incelendi. Elde edilen sonuçlar, Yapay Sinir Ağları algoritmasının gürültülü ve temiz sinyaller üzerinde yüksek empedans arızalarını diğer olaylardan daha yüksek doğrulukla tespit ettiğini, özellik olarak kullanılan standart sapma yaklaşımının ise enerji toplamı yaklaşımından daha iyi performans ortaya koyduğu gösterdi. Bununla birlikte gerek laboratuvar ortamında elde edilmiş yüksek empedans arıza sinyallerinin kullanılmasının gerekse de sinyallere gürültü eklenmesinin, yöntemlerin genel performansında bir düşüşe yol açtığı gözlemlendi.

Konum tespiti öncesi ilk adım olan yüksek empedans arızasının diğer arızalardan ayrıştırılması çalışıldıktan sonra, tez içerisinde yüksek empedans arızalarının noktasal konumunun belirlenmesi konusuna odaklanılmıştır. Yapılan çalışmalara göre, yüksek empedans arızalarını ayırt etmeye kıyasla onların konumunu bulmaya yönelik nispeten daha az çalışma yapılmıştır. Çalışmaların yaklaşık %80'i yüksek empedans arızalarının tam yerini belirlemek için empedans tabanlı yöntemler veya makine öğrenme teknikleri kullanmıştır. Son yıllarda, gerilim ve akım sinyallerinden elde edilen

özellikleri kullanan makine öğrenimi yöntemleri daha düşük hata oranlarına sahip olmaları nedeniyle popülerlik kazanmıştır. Ancak, bu çalışmalardan hiçbiri özellik seçimi için herhangi bir yöntem uygulamamış veya seçtikleri özellikler için bir gerekçe sunmamıştır. Bu nedenle tezde özellik seçimi tekniklerinin yüksek empedans arızalarının konum tespiti üzerindeki etkisinin araştırılması amaçlanmıştır. Bu kapsamda öncelikle yüksek empedans arızalarının akım dalga formları benzetim çalışmaları yardımıyla elde edilmiştir. Daha sonrasında ise akım sinyallere yedi seviye ayrık dalgacık dönüşümü uygulanmış ve her bir farklı dalgacık katsayıları için beş istatistiksel parametre hesaplanarak toplamda kırk adet özellik elde edilmiştir. Bu istatistiksel paremetreler; standart sapma, varyans, çarpıklık, basıklık ve enerji içeriğinin toplamıdır. Daha sonra oluşturulan özellik havuzuna, özellik seçim yöntemleri uygulanmış ve seçilen özellik alt kümeleri ile eğitilen yapay sinir ağlarının sonuçları değerlendirilmiştir. Her bir özellik seçme yönteminden önerilen özellikler yapay sinir ağlarda 250 kez test edilmiş ve değerlendirmelerde tüm sonuçların ortalamaları dikkate alınmıştır. Yapılan değerlendirmelere göre, yapay sinir ağlarının yapısı aynı olduğunda sarmal yöntemler diğer özellik seçimi yapan yöntemlerinden daha iyi performans göstermiştir. Ancak, rastgele orman yönteminin önerdiği özelliklerin kullanıldığı durumlarda, farklı yapay sinir ağları yapılandırmaları için tutarlı bir şekilde daha düşük hata mesafelerine ulaşıldığı ve gürültülü koşullar altında bile test verilerinde daha yüksek doğruluklar elde edildiği görülmüştür. Genel bir değerlendirme yapıldığında ise tüm özellik seçimi yöntemleri, gürültülü koşullarda bile hesaplama karmaşıklığını başarıyla azaltmış ve yöntem doğruluğunu arttırmıştır. Dahası, bu yöntemlerin verilerdeki önemli ve alakasız olarak tanımlanabilecek özellikleri de başarıyla ayrıştırabildiği gözlemlenmiştir.

Mevcut yayınlarda eksik kalan alanları tamamlamayı hedefleyerek, yüksek empedans arızalarının tespiti ve konumlandırılması kapsamında yapılan çalışmaların ardından, tez içerisinde yüksek empedans arızasının meydana geldiği arızalı şebeke dalını ve konumunu doğru bir şekilde hesaplayabilen veriye dayalı yeni bir yüksek empedans arızası konum belirleme yöntemi önerilmiştir. Bu yenilikçi teknik, yürüyen dalga yaklaşımı, Ayrık Dalgacık Dönüşümü ve makine öğrenmesi yöntemi kullanmaktadır. Yayınlarda yaygın olarak kullanılması nedeniyle önerilen metodoloji içerisinde de kullanılan makine yöntemi olarak yapay sinir ağları seçilmiştir. Normalde yürüyen dalga yöntemlerini kullanarak arıza mesafesini doğru bir şekilde tahmin etmek için her yansıyan ve kırılan yürüyen dalgayı tanımlamak gerekmektedir. Ancak bu yöntemin direkt olarak kullanımı özellikle birden fazla kısa dala sahip dağıtım şebeklerinde oldukça zor olabilmektedir. Bu nedenle, arıza konum mesafesini etkili bir şekilde tahmin etmek için önerilen yöntemde yapay sinir ağları girdisi olarak akım sinyali üzerinden kolayca ayırt edilebilecek belirli sayıdaki seçilmiş yürüyen dalga bilgisi kullanılmaktadır. Bu seçim sürecini, yürüyen dalga ve makine öğrenimi tekniklerinin kombinasyonunu ile birlikte kullanarak her yöntemin belirli zayıflıklarının üstesinden gelmek amaçlanmıştır. Önerilen yaklaşım, sağlamlığını ve verimliliğini doğrulamak için farklı arıza başlangıç açıları, makine öğrenme teknikleri, yük değişimleri, aşırı yüklenme koşulları ve dağıtım şebekeleri dikkate alınarak birden fazla senaryo için değerlendirilmiştir. Elde edilen bulgular, önerilen yöntemin hatalı şebeke dalını belirlemede yüksek doğruluk sergilediğini ve arızanın meydana geldiği yerin mesafesinin hesaplanmasında minimum hata oranına sahip olduğu göstermiştir.

1. INTRODUCTION

An electrical arc flash is a serious electrical hazard that poses significant risks. It occurs due to discharge between conductors through the air. This discharge produces very bright ultraviolet light and intense heat during the event. The reason for electrical arcing mainly lies in the technical failure of the equipment or environmental factors that create short circuits. Current flowing through the short circuit determines the difficulties of arc faults. High short circuit current magnitudes can be easily detected by traditional protection equipment. However, low short-circuit currents often go undiagnosed by conventional protection schemes. These faults that do not produce sufficient fault current to detect by protection devices are called high impedance faults (HIF) [1]. Arc events are often related to HIFs that can occur when the live conductor breaks and comes into contact with the ground or touches nearby vegetation. In the case of leaning trees on overhead lines, HIFs may go unnoticed because power delivery to customers remains uninterrupted, and very few noticeable changes occur in the signals. Faults may repeat themselves at different times until they are distinguished, and they may cause fires. The same is also true for broken overhead line conductors. HIFs, caused by broken conductors, create fault currents below the protection thresholds. Therefore, traditional protection systems face difficulties while detecting them. Although they have an undeniable percentage (more than 20%) among all types of faults in the distribution network, 17.5% of HIFs cannot be detected by traditional protection methods [1]-[4]. A study [5] provided evidence supporting this claim through interviews conducted with power line maintenance crews. The study reported that approximately one-third of broken conductor faults remained energized when the maintenance crews arrived at the fault location.

The reasons for not discovering and locating such faults are the random behavior of fault, low amplitude current, and similarities with other events. In most cases, HIFs produce significantly lower currents than the threshold values of protection devices. Figure 1.1 describes the current profile within a standard distribution network. When the figure is investigated, it can be seen that the amplitude of the HIF current is

almost identical to the regular loads in the network. This resemblance challenges its differentiation from standard load conditions at the substation. Therefore, traditional substation protection devices may struggle to identify HIFs if they rely solely on amplitude for detection.

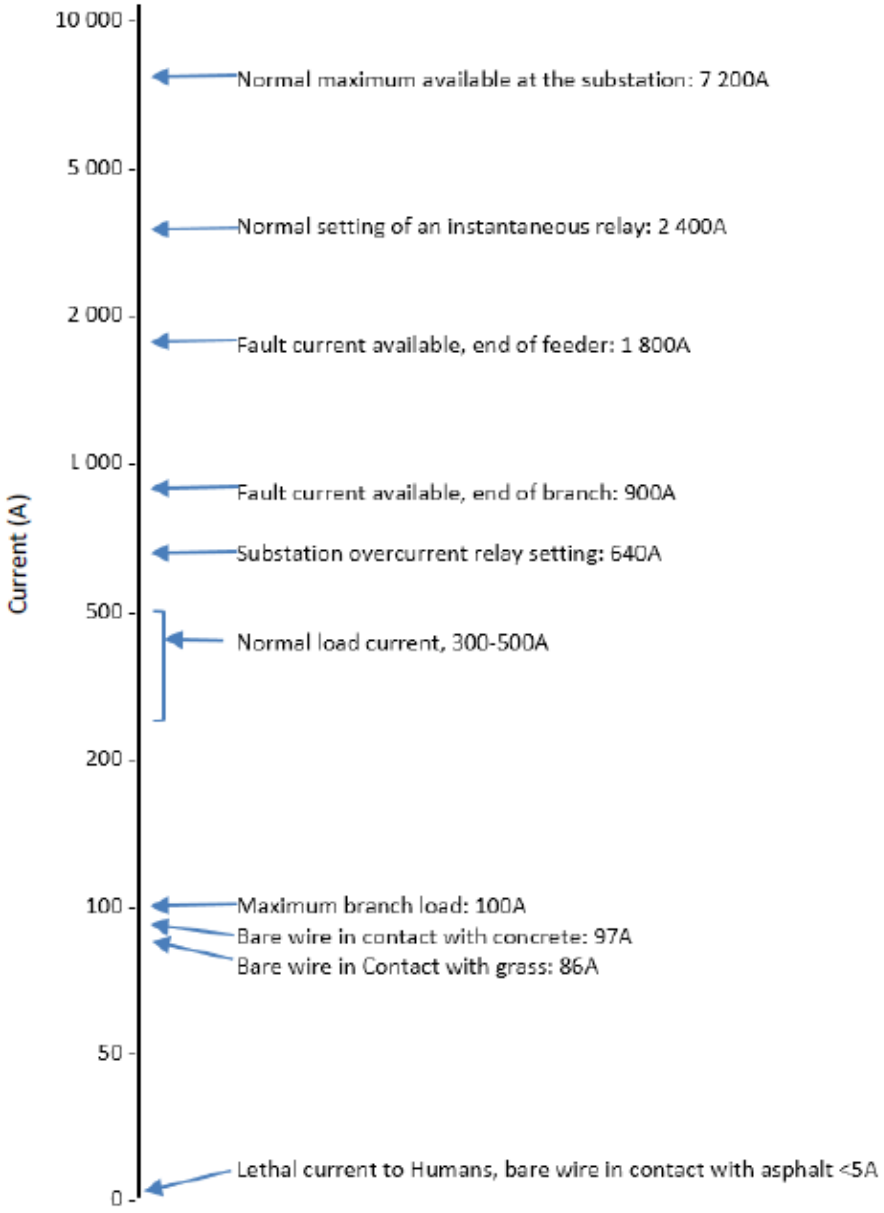


Figure 1.1 : Typical current profile in 12.47 kV network [6].

The HIF currents vary based on the type of contact surface, and their magnitude fluctuates due to weather, moisture, and chemical reactions. This situation creates random behavior characteristics of HIFs that complicate their detection and identification. The similarities of HIF with other events, such as capacitor bank switching and non-linear loads, make it also challenging to detect. The occurring

currents of these events are similar to the characteristics of HIF currents, which can lead to false detection. These findings emphasize the necessity for enhanced detection and localization methodologies to identify HIF events. As a result, researchers have focused on HIFs for over 45 years [7], [8]. Numerous studies have been conducted on HIFs [9], [10]; however, a comprehensive solution for reliably identifying and locating all HIFs is still missing.

1.1 Motivation

HIFs generally do not lead to serious disruptions in the normal functioning of the networks. However, broken live conductors and tree branches touching the live conductors can present a significant risk to the surroundings and people. Arcing, which occurs during HIFs, can start fires. Moreover, HIFs also create power quality problems due to their non-linear characteristics. Therefore, distinguishing and locating HIFs became a critical topic for power utilities. The absence of a universal solution for decades and the growing electrical grids require further studies.

1.2 Research Objectives

The main goal of the thesis is to model, detect, and locate HIFs in overhead lines of distribution networks effectively. The following studies were performed to accomplish these aims.

1.2.1 Understanding the characteristic of HIFs

It's essential to fully understand the fault characteristics of HIFs to detect and locate them accurately. To achieve this, analyzing faulty signals obtained from laboratory tests is crucial. Therefore, high-voltage experiments were conducted at the distribution level in the thesis by considering different types of contact surfaces.

1.2.2 Simulating distribution networks, HIF, and other distribution events.

It is necessary to use various distribution networks to assess the performance of detection and localization methods. These distribution networks should have branches and unbalanced loads to imitate actual distribution network conditions. Moreover, simulating the distribution events having similar characteristics with HIFs is crucial to

ensure the effectiveness of the detection algorithm. Therefore, such events in different networks were simulated in the thesis.

1.2.3 Verify detection algorithm

Although the main aim is to find a feasible solution for localizing HIFs in the distribution network, it is essential to differentiate HIFs from other events first. Therefore, the thesis evaluated the published machine learning-based methods and their offered features. The obtained results are compared with each other, and the effectiveness of methodologies is verified by considering the impact of using actual HIF signals.

1.2.4 HIF localization with feature selection methods

Accurately detecting HIFs is the first step in pinpointing faults and restoring the network. Once HIFs are detected, it is essential to locate these faults to facilitate network recovery. Since it is challenging to identify HIF's location through time domain analysis, extracting distinct features is necessary for HIF localization. According to existing literature, statistical parameters are commonly chosen as features to train the machine learning methods while pinpointing HIFs. Nonetheless, a notable gap in the research is the absence of feature selection strategies and comparative analyses between various feature subsets. Therefore, the impact of different feature selection methods is investigated within the thesis by comparing them based on the error distance.

1.2.5 Develop and validate the localization algorithm

Based on the literature review in the field, a new approach combining the traveling wave method and artificial neural network was proposed. The method's efficacy was evaluated on various test benchmarks with different scenarios and compared with other methods by error.

1.3 Outline of The Thesis

High impedance faults frequently arise in medium voltage networks, particularly in rural regions where overhead lines are uninsulated. They are caused by unintentional

contact with nearby trees or breaking and touching the ground surfaces. This thesis examines arcing HIF in medium-voltage networks, studies existing detection and localization methods and proposes a novel approach for HIF location.

The thesis is organized as follows:

- Chapter 1 provides the motivations behind the thesis and an overview of the thesis objectives.
- Chapter 2 presents a comprehensive review of the literature, focusing on the available HIF detection and localization techniques, as well as the characteristics of HIFs and their electrical models.
- Chapter 3 describes the tests conducted at the ITU Fuat Külünk High Voltage Laboratory. These tests involve a broken overhead line contacting different surface materials and an overhead line interacting with a tree branch. Furthermore, the captured HIF signals are analyzed by considering various voltage levels and surface thicknesses.
- Chapter 4 presents a series of extensive simulations. Two distinct input approaches with two different machine learning methods are examined and compared to verify their accuracy in distinguishing HIFs from other events. Simulations considered three databases that contain actual, modeled, or both types of arcing HIFs to understand their impact on detection.
- In Chapter 5, the effect of feature selection methods on the localization of HIFs was examined by addressing existing gaps in the literature. Subsets of statistical features obtained through various feature selection methods were evaluated and compared, both among themselves and against the outcomes derived from utilizing the entire set of features.
- Finally, in Chapter 6, a combined approach utilizing the traveling wave method and machine learning has been proposed to identify the fault branch and determine the exact distance of the HIF from the source.

2. LITERATURE REVIEW[1]

High impedance faults are an ongoing challenge in power system protection. The power distribution network often experiences arcing HIFs when conductors come into contact with tree branches or ground surfaces. Such faults produce comparatively low currents, making them difficult for power system operators to identify using standard protection devices. Furthermore, an arcing occurring during the HIFs creates a challenging situation for identifying faults. Therefore, a complete understanding of HIFs is essential to develop practical solutions. In this content, topics such as determining characteristics and modeling of HIFs started to get attention, in addition to the detection studies. After the detection studies, localization studies became the point of interest. Although numerous studies have been conducted on HIF, a universal solution to the HIF detection and localization problem has yet to be established. Thus, a comprehensive review of the mentioned subjects by considering some key indices is presented in this chapter. Moreover, a comparative assessment has been performed for both detection and localization methods based on their benefits and drawbacks alongside their efficiency.

2.1 HIF Characteristics from The Literature

Initial studies on HIFs utilized signals obtained from laboratory tests or staged faults in the field. Later, simulation models are introduced with a better understanding of the characteristics of HIFs from these real signals. Identified distinctive characteristics of HIFs can be summarized as [3]-[8], [11], [12], [13]-[16]:

Nonlinearity: Voltage and current signals have a nonlinear relation with each other. This nonlinearity causes harmonics up to 600 Hertz (Hz) for the current and up to 300 Hz for voltage waveforms.

---

[1] : This chapter is based on the following publication: **Baharozu, E.,** Ilhan, S., and Soykan, G. (2023). High impedance fault localization: A comprehensive review, Electric Power Systems Research, 214, Part A.

Asymmetry: There is a difference between the peak value and shape of positive and negative half cycles of HIF current.

Intermittence: HIF current is not steady all the time due to extinction and reignition of the arc, which causes discontinuities in fault current. This arc phenomenon creates high-frequency components in the current waveform as well.

Randomness: Change in impedance leads to changes in magnitude with time.

Build-up and Shoulder: Build-up is the period where fault current grows and reaches its maximum magnitude. On the other hand, the period where the current magnitude remains steady during a few cycles is called the shoulder.

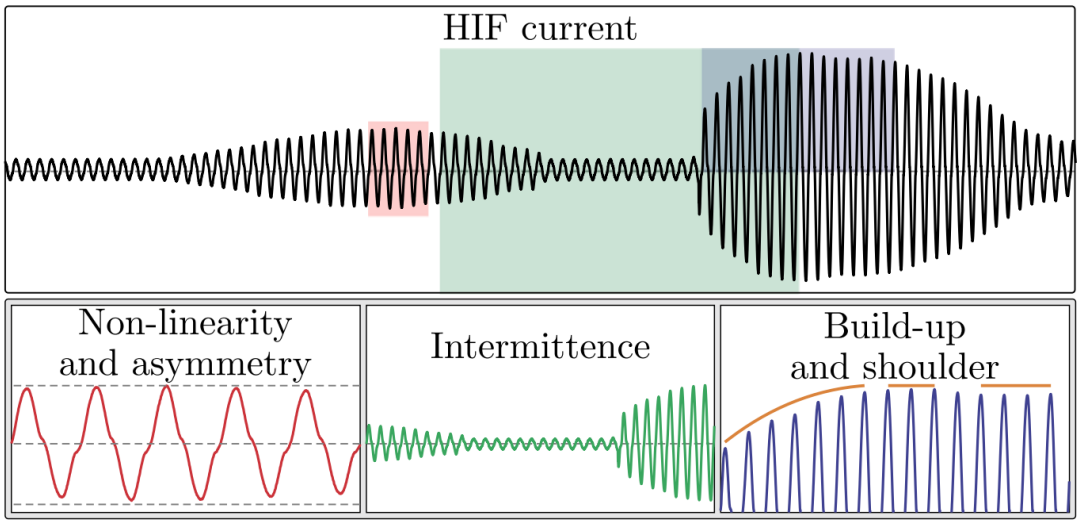


Figure 2.1 : HIF current waveform and distinctive characteristics [17].

2.2 HIF Modeling

Researchers presented various models for arcing high impedance faults from the past by aiming to represent the specific characteristics of HIFs. These models can be categorized into three categories: Differential equation-based dynamic models, electrical elements-based models, and a combination of them.

The dynamic equations are used to calculate the conductance of the arc in the first group of the HIF models [9]-[10]. Such models have existed since Cassie [18] and Mayr [19] were presented. In these models, arc conductivity is given as a first-order differential equation. These equations have been improved and modified in the following decades [20]. The Kizilcay [21] and Johns [22] models are the most popular

formulations. Arcing fault formulation of the Kizilcay model is given below in equation (2.1) and equation (2.2).

$$\frac{dg}{dt} = \frac{G(t) - g(t)}{\tau} \tag{2.1}$$

$$G(t) = \frac{|i(t)|}{(u_0 + R \times |i(t)|) \times Z} \tag{2.2}$$

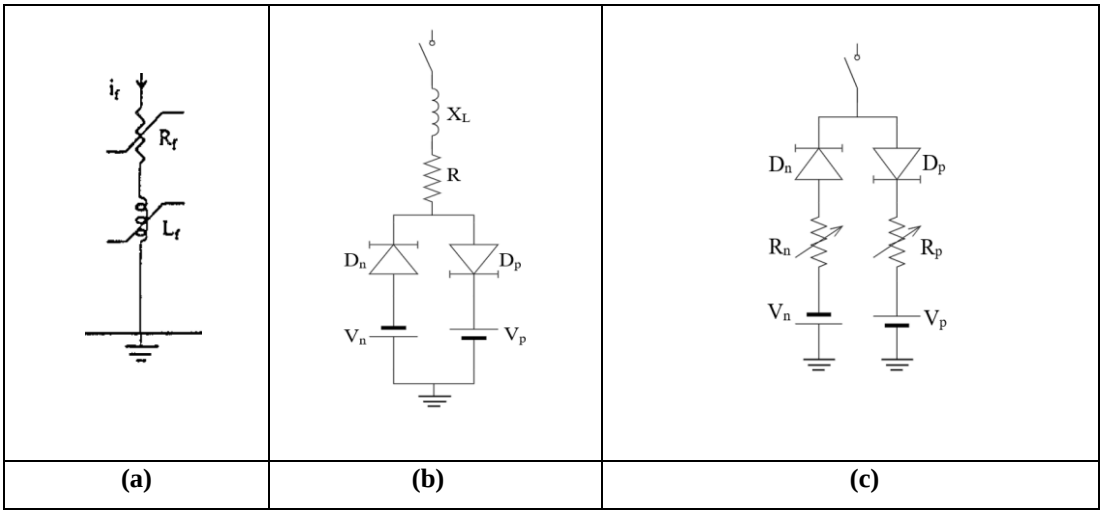
Where g is the instantaneous conductance of the arc, τ is a time constant, G is the stationary conductance of the arc, i(t) is the instantaneous arc current, l is the arc length, R is the resistive component per length of the arc, and u_0 is the constant voltage parameter per length of the arc.

When the Johns model [22] is considered, the formulation of the stationary arc conductance is given in (2.3):

$$G = \frac{|i|}{u_0 \times l} \tag{2.3}$$

Later, a combination of the John and Kizilcay model, known as the universal arc model, was presented [23]. This model has been verified and used in many studies, particularly when arcing faults due to tree leaning are considered.

In the second group, the usage of electrical parameters in the time domain is offered to imitate the behaviors of HIFs. In the beginning, just a fixed resistance was proposed to model HIFs [24]. Almost a decade later, another one [25] offered to use nonlinear impedance as a simple HIF model. A similar approach was also proposed in [26]. Nonlinear and time-dependent resistance and inductance are used as series in this model, as shown in Figure 2.2 (a). A more complicated model that represents the two characteristics of HIF current was presented in [11]. These characteristics were nonlinearity and asymmetry features of HIFs. The model, given in Figure 2.2 (b), consists of constant inductance and resistance series with an antiparallel diode and DC voltage sources. Moreover, field tests in the 13.8 kV system were carried out in the same study to obtain the electrical parameters of the model. In 1993, [27] suggested using two variable fault resistances to imitate the randomness of HIF faults. Thus, three main behaviors of HIFs were considered in this model. An illustration of this model is presented in Figure 2.2 (c).



(a)	(b)	(c)

Figure 2.2 : HIF models in [26] (a), [11] (b), and [27] (c).

On the other hand, [28] offered to use Transient Analysis of Control Systems (TACS) in the HIF model for the first time. The proposed HIF model includes nonlinear impedance, time-varying voltage sources, and a controlled switch with TACS. The purpose of TACS was to establish the arc reignition and extinction. Additionally, voltage sources are taken sawtooth wave shapes instead of DC sources to emulate dynamic V-I characteristics of arc. An illustration of the model is presented in Figure 2.3 (a). Build-up and shoulder phenomena of HIF based on utilizing TACS and two time-varying resistances were added to the HIF model for the first time in [12]. While one resistance in the model was imitating the asymmetry and nonlinearity, the second was imitating the build-up and shoulder phenomena. A representation of the model is given in Figure 2.3 (b). The authors also validated the proposed model with field tests in a 22.9 kV distribution network. Later, [16] improved the model of [12] by adding a switch to simulate the breaking of the conductor (intermittence). Furthermore, field tests were performed in the 13.8 kV network to obtain the electrical characteristics of the model.

In 2003, a new and simplified model based on Emanuel [11] was presented in [29]. Here, two unequal resistances, Rp and Rn, were implemented in parallel, as shown in Figure 2.3 (c). Soon after, [30] offered the six-arc model of [29] in parallel. Besides, a field test in a 20 kV power network was carried out to collect HIF data for the model in the study. Another study [31] suggested adding two inductances, each one series with different unequal resistances, to the model [29] as L_p and L_n.

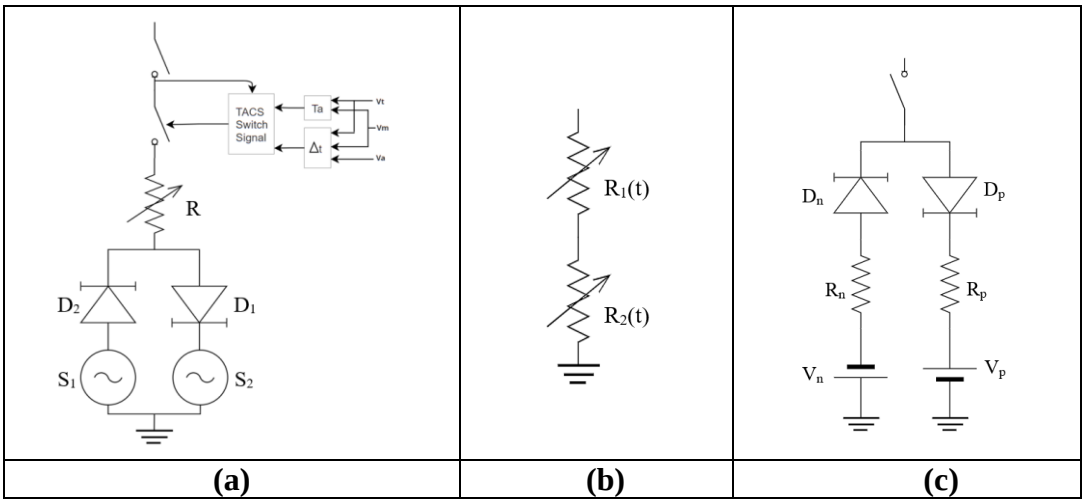


Figure 2.3 : HIF models in [28] (a), [12] (b), and [29] (c).

A novel model is proposed in [32] by combining differential equation-based and electrical elements-based models. To imitate the dynamic characteristics of the HIF, the digital arc model proposed in [21] was applied as nonlinear resistance in the model. DC and AC voltage sources are implemented respectively to obtain asymmetry and arc ignition features. Diodes are used for nonlinearity and control arc ignition in the model. A representation of the model is given in Figure 2.4 (a). The same authors present another model, as shown in Figure 2.4 (c) [9]. The digital arc model proposed in [21] was applied again. However, variable arc voltages in [11] are used instead of diodes and polarizing ramp voltage sources.

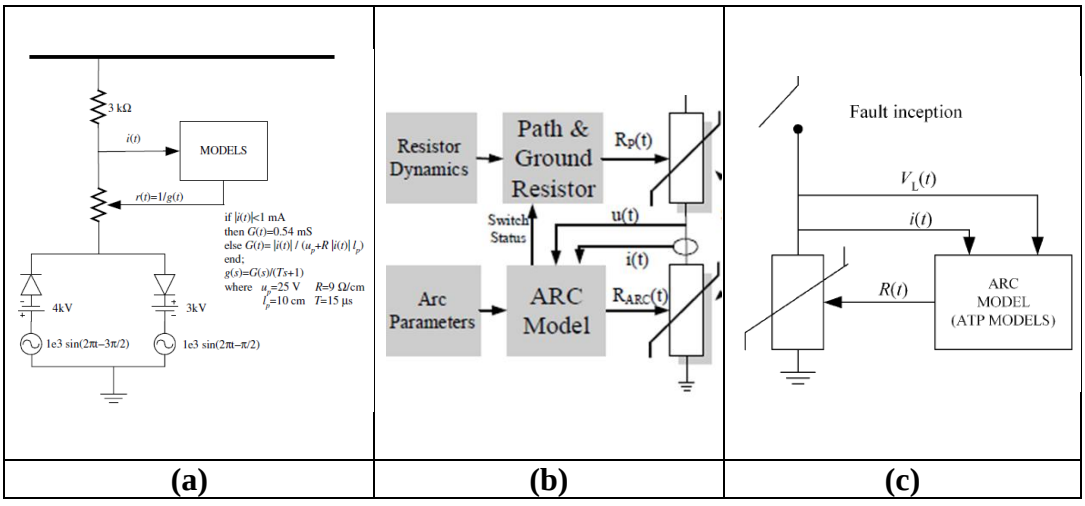


Figure 2.4 : HIF models in [32] (a), [33] (b), and [9] (c).

Another model [33] used two resistances in series. One of them is used for dynamic arc features such as quench and reignition, and the second one is used for varying fault path resistance. While Mayr's equations [19] were chosen for dynamic arc

representation in the first resistance, time-varying resistance is implemented for the second feature. The proposed model is presented in Figure 2.4 (b). A similar approach is proposed in [10] by using two resistances in series. While the dynamic arc model [23] is implemented in one of the resistances, constant resistance is assigned for the second resistance by considering the HIF object as a tree.

2.3 Evaluation Of HIF Detection and Localization Studies in The Literature

Since HIFs cannot be easily detected and located with conventional methods, plenty of studies have been conducted to find a universal solution [34]-[35]. In the rest of this section, published studies in the literature are categorized, evaluated, and discussed in terms of the detection and localization of HIFs.

2.3.1 HIF detection studies

After gaining a better understanding of high impedance faults and their inevitable effects on power systems, researchers focused on improving the detection process. Proposed detection techniques for high impedance or arcing faults can be categorized into four main types: time-domain, frequency-domain, wavelet transform, and machine learning methods. The time-domain algorithms mainly use relays. The earliest study on HIF detection was the proportional relay algorithm [36]. Unbalancing between phases was considered for the HIF detection in the study. Subsequently, the ratio-ground-relay algorithm was proposed based on this consideration [37-38]. Later, another study [39] compared these two studies. Other time-domain researches also have been done and presented in [40-44]. However, such schemas are vulnerable to load variations in the distribution network, and their effectiveness decreases during unbalanced load conditions [45].

To overcome this, researchers move on to new approaches, including feature extractions. These approaches involve using characteristic features of HIFs at different frequency spectrums. One of the first proposed approaches was using high-frequency current components to distinguish the HIFs. For instance, the energy value of the high-frequency signal is correlated with a determined threshold value to detect HIFs [46]. In the study, field tests also were carried out, and the method was verified. Another

study [15] also utilized high frequency and incorporated the randomness feature of arcing.

On the other hand, other detection algorithms mainly used Fourier Transform (FT) to extract the low-order frequency information from the signals. The first study, related to the low-order frequency spectrum, offered to use the third harmonic current [47]. Based on this work, many other studies focus on a harmonic component. One of the earliest experimental studies was performed in [48] to understand harmonic behaviors better during the faults. In the last decades, similar studies have been conducted and presented.

One study [49] claimed that harmonic components such as frequency, magnitude, and phase angle could be beneficial to detect HIF. A study in [50] introduced a new method using sliding. In this approach, the magnitude and phase responses of the current were compared with threshold values to detect HIFs. Another study [51] offered an algorithm that uses a second and third harmonic combination by considering some threshold values. A methodology for detecting HIF that relies on the phase angle of the 3^{rd} harmonic current was proposed in the study [52]. Furthermore, it is declared that the phase angle drops suddenly when the HIF is observed. Thus, it is feasible to identify the fault by examining the phase derivative. Another study [53] used a method called the power spectrum technique. In this method, the values obtained after FFT are squared, and then the calculated average is compared with a specified threshold value to identify the HIF. A study in [54] proposed using harmonic voltage for detection by observing the total harmonic voltage distortion ratio and the third harmonic voltage value at the non-fault phase. Another study used the summation of three-phase harmonic current information instead of one-phase harmonic content. It considered the relation between the 3^{rd}, 5^{th}, and 7^{th} harmonic current for HIF detection [55].

Another feature extraction technique commonly utilized in HIF detection is Wavelet transform (WT). Unlike FT, WT offers enhanced processing capabilities by analyzing the signal within the time-frequency domain [56, 57]. This has made it widely used by researchers in recent decades. WT decomposed to signal into several coefficient levels within various frequency bands. It is categorized into continuous wavelet transforms (CWT) and discrete wavelet transforms (DWT). The majority of HIF detection studies include detail and/or approximation coefficients of DWT [58-65]. Yet, some studies focus on CWT to extract features [66-68].

In [58], current signals are decomposed up to 4 levels with DWT and evaluated in terms of the extinction and reignition of arc phenomenon to distinguish HIF fault. Another study that used both voltage and current signals is presented in [60]. In this study, squares of phase current details are used. The study involved using squares of phase current details and calculating the moving average for each phase. These values were then compared with a predetermined value to identify HIFs. In reference [59], it is proposed to investigate the absolute sum value of the 3rd detail component for phase currents. By comparing these values between phases and feeders, it has been tried to distinguish the arcing fault. Additionally, a study in [61] offered to use and compare 3rd detail components of phase currents. [62] introduced a straightforward detection method using a cutoff value. Here, the cutoff value is compared with the results of an offered simple formula that uses the standard deviation of detail coefficients. The method in [63] suggested utilizing energy changes of 6-level detail coefficients for current and voltage. The study claims that HIFs can be found based on fast-rising energies. [64] proposed to use the energy value of detail coefficients as well. The method involves checking if the calculated energy exceeds the threshold. Moreover, a proposed second hypothesis in the study asserts that HIFs cannot create sags and swell, and the approximation coefficients of voltage signals in terms of energy can be used to discriminate HIFs. In the study [65], the HIF detection algorithm was developed by observing the duration and frequency of the wavelet coefficients exceeding the limit values. Based on such evaluation, it is claimed that HIF and transient events can be distinguished.

Conversely, one of the earliest CWT applications was [66]. Here, the WT magnitude of voltage signals was considered for the discrimination of HIF. In another study [67], CWT was applied to current signals to analyze transformed signals in magnitude, duration, and slope. Then, these values are compared with defined threshold values to identify HIFs. In [9], HIF determination was attempted by observing the phase differences between the CWT coefficients for zero-sequence voltage and current signals. This method was improved and evaluated in [68].

Almost all mentioned feature extraction methods used threshold values to identify HIF events. In these studies, threshold values are mainly determined based on network load characteristics. However, there is a diverse range of loads that change over time in large distribution networks. This situation makes such studies unrealistic [69, 70].

Therefore, authors have started to focus on pattern recognition methods to cope with this problem, such as decision trees (DTs) [71-75], genetic algorithms (GAs) [76-78], neural networks (NNs) [79-92], and support vector machines (SVMs) [93-96].

Frequency-based features are used in [71-73], while wavelet-based feature extraction methods are used in [74-75] to feed decision tree algorithms. [71] proposed to utilize information on root mean square (RMS) current, harmonic magnitudes, and phases in DT to classify HIFs from other faults. The energy of the harmonics is proposed for HIF detection in [72]. Odd, even, and in-between harmonics up to 400 Hz are determined, and their total energies are used as input in the decision tree. [73] proposed a new method for HIF identification. The ensemble decision tree is developed using the amplitude and phase data of harmonic components obtained from current signals with the help of the Extended Kalman Filter. Unlike the previous DT algorithms, [74] presented an algorithm that uses a decision tree together with maximal overlapping discrete wavelet transform (MODWT) to identify HIFs in microgrids. Here, some obtained features by details and approximation coefficients (standard deviation, energy, mean, and entropy) are fed to DT. Similar to [74], the usage of some features such as standard deviation, energy, kurtosis, and skewness are offered in [75] to detect HIFs.

A few studies used heuristic methods such as a genetic algorithm for HIF discrimination. [76] utilized from GA to reduce the size of features and then applied Bayes for fault classification. Here, wavelet decomposition is implemented to get features from signals. Contrary to other studies, isolator leakage current is considered a different fault type in this paper. The same authors presented a different method in [77]. This time, they offered a fuzzy interference system for fault classification instead of Bayes. [78] proposed a real-coded genetic algorithm that utilizes current harmonics and phase angle shifts. The occurrence of the 3^{rd} harmonic, 5^{th} harmonic, and phase shift at the 3^{rd} harmonic are considered for HIF detection.

On the other hand, most studies utilize neural networks for pattern recognition in HIF detection methods. For example, in [79], variations of discrete wavelet coefficients are employed as input in an ANN to identify HIF events. In [80], the energy contents of the first three levels of details are chosen as features to train a neural network classifier. [81] proposed to use the standard deviation of detail and approximation coefficient of specific sub-bands (D1-D7 and A7) within ANN to detect HIFs. Another method in

[82] proposed to train a neural network with a variance of each wavelet coefficient. [83] offered to use features of coefficients such as mean, standard deviation, skewness, and kurtosis. These features are estimated after DWT decomposition and used as input in NN to identify HIFs. [84] presented an HIF detection method that uses absolute summation of D1 detail coefficients during a single cycle period to feed ANN.

[85] proposed to use RMS ratios of voltage and current signals obtained after DWT decomposition in several sub-bands. Here, these ratios are used to train ANN for HIF detection. A new method that uses a convolutional neural network (CNN) and continuous wavelet transform was presented in [86]. CWT is employed to get time-frequency-based grayscale images of currents. Then, these images are used to feed the CNN classifier for fault detection. A similar approach was proposed in [87], where two-dimensional maps derived from wavelet packet entropy analysis are input into the CNN to distinguish HIF events. In [88], a modified wavelet transform is employed to extract features of HIF, followed by training a CNN algorithm using these features. Authors in [89] developed and tested a HIF detector using real field data. The detector's method utilized both the wavelet transform and a neural network. Extracted features from zero sequences of current feed to a neural network in the device.

Although most studies utilized wavelet-based features, some used frequency-based features to train NN. One of these studies is [29]. Low-frequency voltage and current data obtained by finite impulse response filter are used here. Then, these low-frequency waveforms feed a neural network classifier for HIF detection. Another study [90] proposed to use the 3^{rd} and 5^{th} harmonic vectors of current and voltages to train an artificial neural network algorithm. In a parallel way, the study in [91] presented a HIF identification method that uses harmonic and sequence components currents to feed ANN. [92] is also proposed a HIF detection schema that uses low-order harmonic currents as neural network inputs.

All the studies that took advantage of SVM algorithms as HIF classifiers employed wavelet transform to derive features from signals. The HIF identification method based on the energy of DWT details and approximation coefficients as input in SVM is presented in [93]. Here, the proportion of each 8-level detail and 8th approximation coefficients energy to the total energy of the current signal fed to SVM. [94] presented a similar approach by utilizing DWT for feature extraction and SVM for event classification. Moreover, different mother wavelets such as Daubechies 4 (db4),

Coiflet (Coif5), and Symlet (sym5) are compared and tested in the study. Another study [95] trained SVM with different data sets as normalized reactive power, RMS of detail coefficients, and standard deviation of decomposition. [96] proposed to optimize SVM with particle swarm optimization (PSO). The proposed method was also compared with another SVM method that utilizes FFT for feature extraction instead of DWT. Based on the results, it is claimed that the proposed method is superior to the traditional one.

The most promising or complicated ones among HIF detection studies are given in Table 2.1 to compare studies more clearly. It provides information such as measured signals, test systems, detection methods, cases, and test implementation conditions. Additionally, considerations of nonlinear load, unbalancing, and noises are expressed in the table. It is clear that [55] fulfills most of the criteria. However, [55] did not study large distribution networks, HIF model verification, and different networks.

Table 2.1 : HIF detection papers.

Ref	Year	Measurement	Detection Method	Test System	Nonlinear Load	Unbalance Loading	Noise
[51]	2016	current	Frequency -Based	IEEE 13	Yes	Yes	No
[52]	2016	current	Frequency -Based	90 Buses (Real)	Yes	No	No
[55]	2018	current	Frequency -Based	IEEE 13	Yes	Yes	Yes
[64]	2017	voltage and current	Wavelet	90 Buses (Real)	No	No	Yes
[73]	2012	current	Decision Tree	14 Buses (meshed)	Yes	No	Yes
[75]	2018	current	Decision Tree	India - Real Basic Fider	No	Yes	Yes
[92]	2020	current	ANN	IEEE 34	Yes	Yes	Yes
[95]	2018	voltage and current	SVM	202 Buses (Real Data)	Yes	Yes	No

2.3.1.1 The popularity of HIF detection techniques

For detection studies, numerous articles can be found in the literature since a universal solution is still not found. Yet, in this thesis, mainly the latest 61 studies are evaluated. One key finding from the reviewed papers is the growing interest in this field, particularly over the past decade. The number of published papers in the literature continues to grow each year. Figure 2.5 provides a quantitative representation of the

reviewed papers over the years. The topic's significance has notably increased in the last decade, especially after the HIF characteristics and models were revealed clearly.

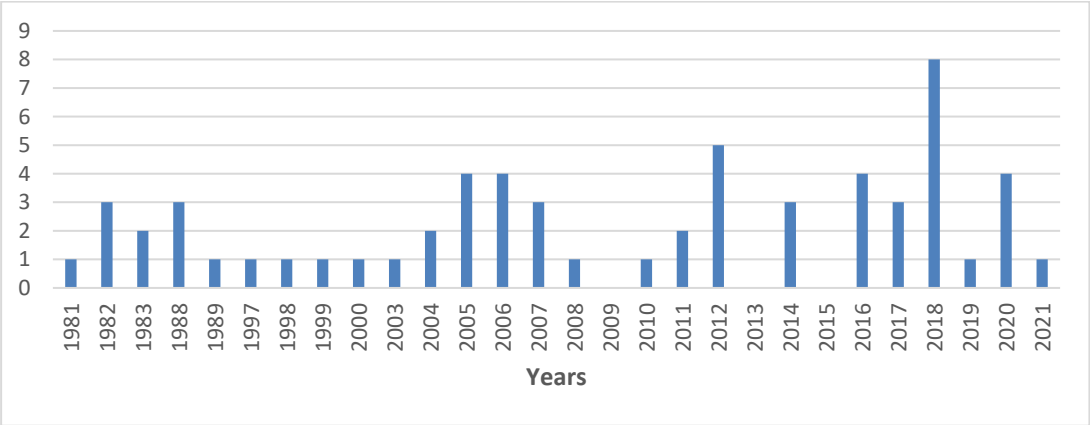


Figure 2.5 : Number of published detection papers by year.

2.3.1.2 Discussion of proposed HIF localization techniques in the literature

After reviewing the studies in the literature, it can be deduced that there is still no universal method available for detecting HIFs. Moreover, there are numerous gaps in the existing studies. Before closer evaluation of the papers on HIF detection, it is valuable to categorize HIF detection studies to understand the tendency of researchers. The percentage of proposed methods as a pie chart is given below in Figure 2.6. It can be seen that machine learning methods are superior to others due to their high accuracy. On the other hand, there is no distinct difference between other studies in terms of preferability.

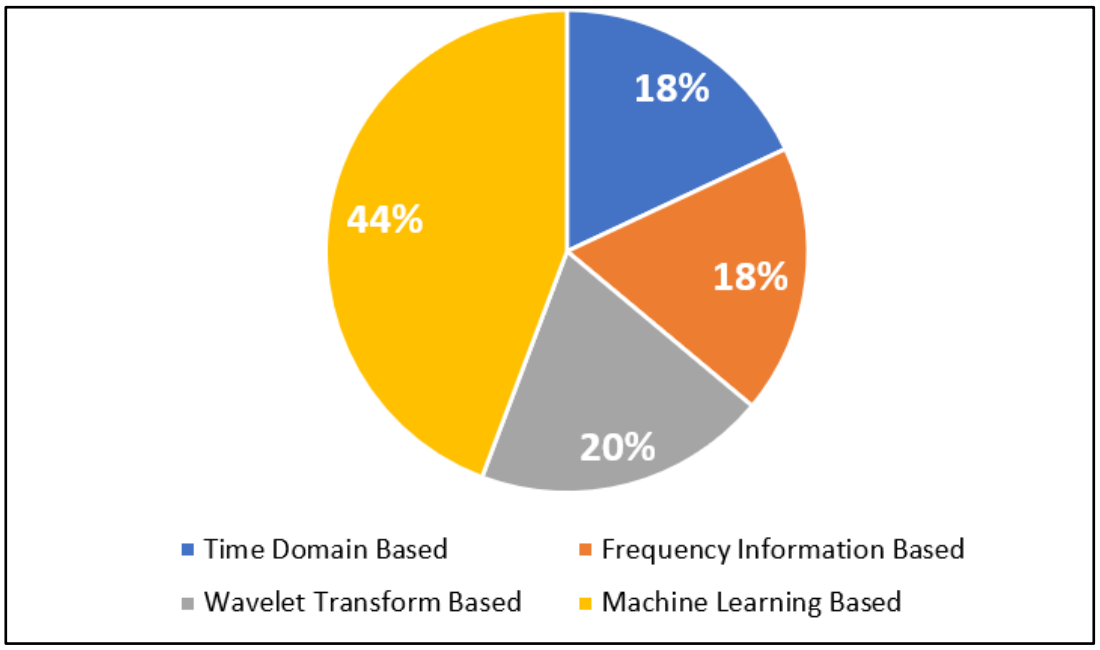


Figure 2.6 : Number of proposed methods for detection of HIFs.

Once detection papers for distribution networks are investigated in terms of their test networks, it's evident that most studies tested their methods on a simple network. Only about one in nine studies utilized larger networks. Moreover, almost none of the methods have been tested in different networks to assess their effectiveness more accurately.

In the investigation of used HIF models in various proposed methods, it was found that only 10 out of 61 reviewed detection papers conducted laboratory or field tests to utilize staged fault signals within their simulations. Although most detection studies consider nonlinear arcing HIF models, the claimed precision in these studies relies heavily on the electrical parameters of the model, which can vary in real-life scenarios. Therefore, the reliability and effectiveness of these approaches in real-world applications remain uncertain.

Additionally, loading variation, a crucial factor in distribution networks, has been overlooked in numerous studies. Different loading conditions can significantly impact the accuracy of detection methods, but only eight detection papers analyzed various loading scenarios. Furthermore, unbalanced loading is not considered in most studies. Similarly, the influence of noise has been largely neglected in the studies. Even though noise is an inevitable aspect of real power systems when measuring voltage and current signals, only eight detection papers examined how noise influences the performance of their proposed algorithms. The characteristic behavior of HIFs is similar to the non-linear load. To prevent false detection, distinguishing between these two signals should be assessed. However, only a few studies considered the non-linear load in their test system.

2.3.2 HIF localization studies

Unlike the HIF detection studies, few studies are available in the literature related to HIF localization. In the available studies, while some of them [97]-[118] focused on finding faulty sections as a zone, others [122]-[163] concentrated on finding the exact location of HIFs. Studies not focused on pinpointing the precise location of high impedance faults can be categorized as wavelet transform-based techniques, voltage changes-based methods, machine learning methods, harmonic information-based methods, zero sequence current-based methods, and others.

In one of the DWT-based studies [97], the faulty section is determined by evaluating the polarity of frequency band power. Here, DWT detail coefficients of the measured residual current and voltage are multiplied, and then calculations are summed during two periods. Afterward, the faulty section is identified as behind or ahead of the node based on the polarity of summation. On the other hand, [98] proposed to compare the summation of the first three levels of detail coefficients decomposed by DWT with the pre-generated database by using the average of absolute difference to find the faulty area. The study is extended in [99] by examining different sampling rates, window sizes, and mother wavelets. [100] also proposed a method using DWT to reduce the search area of HIFs. First, voltages, measured at various locations, are transformed into detail coefficients. Then, their estimated energies are compared to find search areas based on buses having the highest three energy values. [101] presented a new algorithm utilizing Wavelet Packet Transform (WPT). Coefficients of current signals are used to decide a faulty section by distinguishing internal and external faults. Differences between the sum of current coefficient values for each phase are averaged and compared with the predefined threshold value. Based on the comparison, the faulty section was determined. Another WPT-based method is proposed in [102]. The proposed method first applies WPT to measured time-synchronized voltage signals and calculates maximum absolute values of detail coefficients from computed model voltage for all buses in a system. After that, calculated values for each bus are compared with others to determine the faulted area.

As a different type of method, changes in voltage signals are considered. For instance, [103] claimed that faulty bus and fault impedance might be found by observing the characteristic voltage sags. Thus, the authors offered to compare measured and simulated voltage profiles obtained from a known network and a fault to determine the fault location and its associated impedance. An alternative method based on voltage measurement is presented in [104]. The positive sequence voltage drop value is estimated with the help of located sensors at the end of each feeder in the method. The proposed methodology finds the sensors placed below the fault location and identifies them as active by checking the voltage drop value. Then, from each active sensor to the source, nodes are listed to find possible routes. Finally, the last two downstream nodes on common nodes are determined to locate the faulty section. A new method to estimate a faulty section using unbalanced factors, calculated from smart meters

located at the secondary side of transformers, is presented in [105]. Yet, with the increment of fault impedance, the method's effectiveness is negatively affected. [106] introduced a method called Advanced Distortion Detection Technique (ADDT). Distortions of samples in voltage signals are computed as a percentage and ranked accordingly in the approach. Searching the area between measurement points with the highest risk level is offered to locate the faulty section.

As an alternative to the previously mentioned two methods, a few studies analyzed harmonic data to estimate the rough area of the fault. The earliest studies [107] utilized the recloser/sectionalizer technology of the era to use harmonic information. Communicated sectionalizers identify the faulted line by providing information on whether the fault is recognized. [108] proposed to use also harmonic information to find the faulty zone. Second harmonic indexes are computed using voltage signals measured by smart meters. Then, the smart meter with the highest value and adjacent ones are defined as possible faulty zones.

Although many machine learning-based methods addressed locating the exact point of fault, four papers used them to reduce the search area of the fault. A technique based on a support vector machine is proposed in [109]. In this study, the first three details, decomposed from current signals using DWT, feed the SVM to identify the faulty section. Furthermore, the various number of measurement points and their placement based on the modified *k*-means algorithm are investigated in the study. The algorithm called Self Organizing Mapping Neural Network (SOMN) is presented for the same network in [110]. The wavelet entropies of the measured fault and neutral currents trained the SOMN to find the faulty line. Another study [111] used an ANN to find the bus with HIF and to estimate the magnitude of the fault resistance. The study used different electrical parameters such as voltage magnitudes, angles of voltages, active and reactive power, fault current, and fault angles to train ANN. Moreover, it is emphasized that distance error in the study lies between +/- one km, which reduces the required effort to pinpoint the exact fault location by 49%. Another ANN-based method was proposed in [112]. Installed PMUs at branches captured the voltage and current signal to train ANN. The fault location is estimated in 25% increments between buses.

Another approach to finding a faulty section is zero sequence current (ZSC) based methods. One of these methods [113] is called Zero-Sequence Non-synchronized

Protection. This method uses non-synchronized RMS zero-sequence currents to determine the fault section. The difference between two adjacent measurements of zero-sequence currents' phasor magnitudes is compared with the predetermined threshold value to distinguish the faulty area. The proposed method is tested and verified with a Real-Time Digital Simulator. Later, the same authors extended the proposed method in [114] by considering the probable measurement error in the comparison step. Moreover, they tested it in IEEE 37 and IEEE 123 bus test systems with a more comprehensive approach. A new zero sequence-based method for HIF location in a resonant grounded network is presented in [115]. First, the concept of zero-sequence current, which changes between healthy and faulty feeders, is proved in theory in the study. Then, a similar idea of previous studies considering the phase of 3rd harmonic ZSC is applied. Phase differences between selected and remaining feeders are calculated and compared with the threshold value. If the differences between the selected feeder and others are satisfied, that feeder is accepted as the faulty feeder. The proposed method is tested with real experimental cases for different HIF scenarios. Another zero-sequence current-based method is presented in [116]. Zero sequence currents are measured from each feeder. Then, the faulted feeder is determined based on the integrated inner product value (IIPV). This value is calculated by taking the cumulative sum of the inner product of zero sequence currents of each feeder and neutral point. Two pieces of information from IIPV, sign and amplitude, are compared with each feeder to decide the faulty feeder. When the value of any feeder is higher than the others and has the opposite sign, it is claimed that a fault occurred at that feeder. Another approach, combining the inner product with ZSC, is proposed in [117]. A faulty feeder is found by comparing the summation of the elements of each row of the dot product matrix in the study. An improved Hausdorff distance algorithm creates the matrix with the calculated frequency information of zero sequence currents. Both simulation and actual HIF data from the field are used to verify the method. A novel approach based on declining periodic components of ZSC for resonant grounding systems is offered in [118]. The relative phase differences between each node's zero-sequence current's periodic components and the fault resistance are used to define the faulted section.

Besides these methods, [119] presented a different approach to distinguishing a faulted feeder from others by comparing the magnitude, integral value, and spectrum of

complex sequence currents at feeders. Moreover, the complex sequence is compared with the zero-sequence and outperformed in terms of sensitivity in the study. The authors of [119] proposed another method in [120]. Power spectrums of the modal component of the instruct mode function, obtained by decomposing the zero sequence of current and voltage via Ensemble Empirical Mode Decomposition (EEMD), are compared to identify the faulted line. The study identified the faulted bus and the feeder based on maximum voltage and current spectrum. The main details of reviewed papers that focus on determining a faulty section for HIFs are given in Table 2.2.

Table 2.2 : Details of faulty section papers for HIF localization.

Ref.	Year	Measurement	Locating Method	Test System	Unbalance Loading	Load Variation	Noise	Operating Voltage [kV]	Field Test	HIF Model
[97]	2008	Voltage	DWT	14 buses	No	No	No	20	No	[23]
[98]	2012	Voltage	DWT	38 buses	No	No	No	11	No	[24]
[99]	2016	Voltage	DWT	38 buses	No	No	No	11	No	[24]
[100]	2013	Voltage	DWT	90 buses	No	No	No	13.8	No	[12]
[101]	2015	Current	WPT	2 Buses	No	No	No	NS	No	[30]
[102]	2020	Voltage	WPT	3 Buses-5 Buses	No	Yes	Yes	345	No	[24]
[103]	2014	Voltage	Voltage	13 buses	No	No	Yes	13.8	No	[24]
[104]	2005	Voltage	Voltage	28 buses	No	No	No	20	No	[24]
[105]	2018	Voltage	Voltage	7 buses	Yes	No	No	13.8	No	[24]
[106]	2020	Voltage	Voltage	IEEE 13 & IEEE 34	Yes	No	No	4.16 - 24.9*	No	Modified [27] by adding arresters
[107]	2005	Current	Harmonic	5 buses	No	No	No	13.8	No	[12]
[108]	2019	Voltage	Harmonic	IEEE 39	No	Yes	No	345	No	[27]
[109]	2015	Current	SVM	18 buses	No	Yes	No	12.5	No	[11]
[110]	2017	Current	SOMN	18 buses	No	Yes	No	12.5	No	[11]
[111]	2015	Voltage, Current, Power, etc.	ANN	45 buses	No	No	No	13.8	No	[24]
[112]	2018	Voltage & Current	ANN	15 buses	No	Yes	No	NS	No	[24]
[113]	2016	Current	ZSC	11 Buses	Yes	No	No	13.8	No	[24]
[114]	2021	Current	ZSC	IEEE 37 & IEEE 123	Yes	Yes	No	4.8-4.16*	No	[24]
[115]	2022	Current	ZSC	Field and 5 Buses	No	Yes	No	10	Yes	[121]
[116]	2022	Current	ZSC	Modified IEEE13- 34	Yes	No	Yes	4.16 - 24.9*	No	[24] and [27]
[117]	2022	Current	ZSC	Field and 7 Buses	No	No	Yes	10	Yes	[24]
[118]	2020	Voltage & Current	ZSC	9 buses	No	No	Yes	10	No	[118]
[119]	2017	Current	Complex sequence	10 buses	No	No	No	10	No	[29]
[120]	2017	Voltage & Current	EEMD	10 buses	No	No	No	10	No	[29]

*Voltage level for second modeled network
NS is for Not Stated

When considering only the studies that precisely localize fault distance within the distribution and transmission network, they can be categorized into four groups. These are parameter estimation-based methods [122]-[138], traveling wave-based methods [142]-[147], machine learning-based methods [148]-[161], and others [162]-[163].

Most of the HIF localization studies in the literature have primarily used parameter estimation (impedance-based) methods because of their simplicity. For instance, [122] utilized the Least Error Squares approach to estimate the location of HIF in the transmission system using the time domain data. In contrast, studies in [123] and [124] employed the same approach in the frequency domain by utilizing synchronized measurements. Similarly, [125] presented a method in the frequency domain to estimate fault distance, where the system equations were calculated using the weighted least squares technique. Later, [126] aimed to enhance the method presented in [125] by analyzing the statistical behavior of the estimates. This approach improved the localization of the faulty segment of the distribution network.

The study in [127] also employed the least squares method in the time domain to estimate fault distance. This approach involves taking voltage and current measurements from both ends of the transmission line, along with the line parameters, for calculations. Results indicated that the error ratio was less than 10% in over 80% of the test cases. Another study [128] suggested using the Gauss-Newton method as an alternative to the least squares method to minimize distance errors. A comparison of this method with the presented method in [127] revealed that the Gauss-Newton method was superior, yielding errors of less than 0.1% in over 90% of the test cases. In [129], the behaviors of HIFs are modeled as a mathematical model in the spectral domain, and the Weighted Least Squares (WLS) method is proposed to calculate fault location. The proposed method is validated using the IEEE 13 bus test system. Another WLS-based method in the time domain, that utilizes synchronized voltage measurements was also presented in [130].

On the other hand, [131] proposed a new procedure based on comparing the pre-fault impedance with the estimated impedance regarding both magnitude and angle to locate HIFs. [132] proposed an HIF fault locater method that uses impedance loop calculation by employing real and imaginary parts of the equation. In [133], researchers utilized linear least square and steepest descent-based iterative search techniques to calculate the fault distance of HIFs within the distribution network in the time domain. Another

study [134] gathered third harmonic voltage and current information through synchronized harmonic phasors to estimate the location of HIFs in terms of impedance value. This method was validated using the IEEE 33 bus distribution test system. The proposed approach in [135] also utilizes an IEEE distribution-level test benchmark to identify fault locations. It compares the measured voltage with the estimated voltage by considering faults at different locations within the network. A distinctive study to find the location of HIFs in a distribution network with lateral lines is presented in [136]. The frequency components of the measured voltage and current signals are employed to calculate the fault distance. On the other hand, a traveling wave is utilized to determine the fault path based on the wave's characteristic frequency. Another novel approach is proposed in [137]. The proposed method first identifies whether the fault is internal or external by using positive sequence admittance of both ends of the line. Then, it calculates the exact fault location using synchronized positive and negative sequence voltage and current phasors through mathematical formulas. The suggested method has been tested on a compensated transmission network with two buses. The maximum error is calculated as 0.041%, while fault impedance varies between 0–600Ω. In a related paper [138], the same authors focused solely on synchronized positive sequence voltage and current phasors. The method was also evaluated using a similar transmission and the IEEE 9 buses test system. In this case, the impedance of the fault was chosen to be 300Ω and 100Ω, respectively. The observed error in fault location was approximately 1.2%.

Although the error percentage of impedance-based methods relies on the accuracy of fundamental components, many possible factors, such as the accuracy of current/voltage transformers, mutual impedance, the impedance of fault, and line impedances, are avoided in traveling wave methods [139-141]. Therefore, some authors pay particular attention to traveling wave methods for pinpointing the location of HIFs. All the studies using the traveling wave method except [142] and [143] are executed for the transmission network. [142] proposed a traveling wave method to find HIF location in a compensated small distribution network by grounding the neutral with a controlled thyristor. On the other hand, [143] proposed a method for HIF location using a complex space-phasor combined with the Hilbert–Huang Transform (HHT). The researchers claim that the complex space-phasor is superior to other transformations, such as the Karrenbauer and Clarke transformations, and with

combination of it with the HHT provides higher accuracy for locating faults. To determine the exact time and extract traveling wave points, synchronized two three-phase voltage measurements from both ends of the network are used. These measurements are transformed into the absolute values of complex space-phasor and processed using the HHT. After that, the arrival times of the traveling waves at each endpoint are calculated based on traveling wave theory. These arrival times are then employed to determine the location of the fault accurately. Simulation results indicated that the HIF location was calculated with an error of 100.3 meters. However, it is crucial to note that the study was conducted with only one HIF location in a simple distribution network.

One of the other studies [144] presented a traveling wave method based on inceptions obtained through wavelet transform on transient voltage data. Due to utilizing two-terminal data in the proposed method, the offered approach needs communication and synchronization. The study in [145] also applied wavelet transform to identify fault waves utilizing current signals from two terminals rather than voltage signals. [146] presented a traveling-wave-based fault location method for a hybrid transmission network. The algorithm used the measured three-phase voltages via optical sensors. The summation of aerial modes is used to calculate the arrival times of traveling waves, employing two sliding windows and curve fitting to ensure the proposed algorithm performs effectively with all fault types. Polarity changes of traveling waves are calculated for each possible fault section, including the first and second halves of the overhead line and underground cable in the studied network. Next, the faulty area is identified within four possible sections by analyzing the polarity changes of three consecutive traveling waves. Once the section is selected, the exact location of the fault is estimated using the time differences between the first and second traveling waves. Contrary to previous traveling wave studies, the proposed method in [147] has been tested in the field rather than through simulations. Furthermore, a commercial device that provides high and low-frequency wave information was employed to detect fault waves.

As an alternative to impedance and traveling wave methods, machine learning algorithms are implemented in some studies to identify HIFs' locations. Most of the proposed machine learning studies in the literature were NN-based methods. One of the studies [148] used phase currents and energy of their coefficients obtained by

wavelet transform as inputs for ANN. Another study [149] used terminal current and voltage waveforms to train ANN for HIF location. [150] proposed a method utilizing two ANNs. One is used for zone location, and the other is used for estimating the exact location of HIF. A similar approach is proposed in [151] as well. This method uses two different ANN algorithms to determine the faulted area and find the exact fault location. The first features of measured three-phase currents and voltages are extracted with a DWT and then used to feed ANN. The proposed method was tested in a double-circuit transmission line and verified with 10,000 faulted cases with a low estimation error of 0.001%. [152] presents a method that uses a neural network whose activation function is wavelet. It is called Wavelet Neural Network (WNN). Furthermore, DWT is used in this study to eliminate noises from signals before estimating the phasor magnitudes of 3 phases that train to WNN. Similarly, [153] used phasor magnitudes of voltage and current signals with the 3^{rd} harmonic value of current as input for an ANN. Conversely, the study [154] employed an online ANN to determine and identify unknown parameters related to current and voltage equations rather than calculating the fault distance. This methodology was compared with the approach outlined in the study [133], as both studies are fundamentally anchored in parameter estimation techniques. Another study, presented in [155], also used online trained ANN, where the fault location was treated as a weight of the ANN. The aim was to identify fault locations by using current measurements as input and measured voltage as output in the ANN. Additionally, [156] offered to use harmonic components as input for the ANN. Fundamental and harmonic components, derived from three-phase voltages and currents through Fourier Transform, are utilized to identify the precise location of HIF in the study. Besides ANN methods, one study [157] suggested using an Extreme Gradient Boosting (XGBoost) algorithm. In this approach, three-phase zero-sequence voltage and current signals are processed using the Teager-Kaiser Energy Operator (TKEO). The resulting coefficients from the TKEO are summed to form a feature vector, which is then fed into four XGBoost modules to locate various types of faults. In most cases, the error in locating HIFs is found between 0 to 2 kilometers.

Other machine learning methods have also been proposed in addition to the traditional neural network-based approaches. The Adaptive Neuro-Fuzzy Inference System (ANFIS), which combines the advantages of ANN and fuzzy logic, is offered in [158] to identify the locations of HIFs. In this study, the input to the ANFIS consists of the

amplitudes of three-phase voltage and current. Another study [159] suggests employing specific orthogonal components as inputs in fuzzy inference systems. Based on the simulation results, it is stated that fault distance can be estimated with a 20% error margin in approximately 90% of cases. A different approach utilizing Gaussian Process Regression (GPR) is proposed in [160] to find the location of HIFs. Some features are extracted in the study by applying DWT to current signals, such as mean, median, and standard deviation. Afterward, these features are used to train GPR. One study [161] introduces a method including SVM to find the exact location of HIF. Although details about methods are not explained explanatorily in the article, it is claimed that the method's accuracy is promising.

Although several machine learning-based methods have been proposed in the literature, it's important to remember that their performance relies heavily on the specific network being tested. Therefore, if the network or its configuration changes, the machine learning algorithm must be retrained using a diverse set of test cases.

In the literature, some studies have proposed innovative methods for locating HIFs. For instance, one study [162] combines an SVM and the traveling wave method to find a fault location in the transmission network consisting of an overhead line and an underground cable. SVM is utilized to identify the HIF location -whether it is a cable or overhead line- and determine which half of the section is faulty. The inputs for the SVM include voltage wavelet energies obtained from DWT and transient currents. After the faulty area is determined with SVM, the traveling wave method is employed to estimate the fault distance precisely.

Another study [163] introduces a new fault location method for compensated networks along with a device designed to implement this method. This approach uses pre-calculated data and current injection at the neutral point. Zero-sequence impedance values for various locations across different feeders and fault resistances are estimated based on the measured signal frequency. The method calculates the highest absolute negative value of the zero-sequence powers to identify the faulty feeder using the injected current. The fault location is identified by comparing the pre-calculated zero-sequence impedances with the measured values. Field tests were conducted with two Hungarian distribution system operators to assess the proposed method. The findings show that the methodology identifies the exact fault distance. However, it often presents multiple potential locations for the fault.

A comparison table summarizing HIF location studies focused solely on accurately pinpointing HIFs is provided in Table 2.3.

Table 2.3 : Details of pinpointing papers for HIF localization.

Ref.	Year	Measurement	Locating Method	Modeled Test System	Unbalance Loading	Load Variation	Noise	Operating Voltage [kV]	Field Test	HIF Model
[122]	2000	Both**	Impedance	2 Buses	No	No	Yes	400 - 110*	Yes	[122]
[123]	2014	Both**	Impedance	2 Buses	No	Yes	No	230	No	[123]
[124]	2016	Both**	Impedance	2 Buses	No	Yes	No	230	No	[124]
[125]	2017	Both**	Impedance	IEEE 13	Yes	No	No	4.16 - 23.1*	Yes	[24]
[126]	2020	Both**	Impedance	IEEE 13 & IEEE 34	Yes	No	Yes	4.16 - 24.9*	No	[24]
[127]	2020	Both**	Impedance	6 Buses	No	No	Yes	229.174	No	[124]
[128]	2021	Both**	Impedance	4 Buses	No	No	Yes	229.174	No	[124]
[129]	2019	Both**	Impedance	IEEE 13	Yes	Yes	No	4.16	No	[11]
[130]	2015	Voltage	Impedance	9 and 22 Buses: Part of IEEE 118	No	No	No	345	No	[24]
[131]	2012	Both**	Impedance	IEEE 13	Yes	Yes	No	4.16	No	[24]
[132]	2015	Both**	Impedance	8 Buses	No	Yes	No	NA	No	[24]
[133]	2016	Both**	Impedance	IEEE 13	Yes	No	Yes	4.16	No	[24]
[134]	2017	Both**	Impedance	IEEE 33	Yes	No	No	12.66	No	[29]
[135]	2019	Both**	Impedance	IEEE 34	Yes	Yes	Yes	24.9	No	[24]
[136]	2014	Both**	Impedance & Traveling wave	3475 Buses	Yes	Yes	No	13.7	No	[24]
[137]	2020	Both**	Impedance	2 Buses	No	Yes	No	400	No	[24]
[138]	2021	Both**	Impedance	2 Buses- IEEE 9	No	Yes	No	400	No	[24]
[142]	2015	Voltage	Traveling wave	10 Buses	No	No	Yes	20	No	[24]
[143]	2012	Voltage	Traveling wave	2 Buses	No	No	No	10	No	[21]
[144]	2011	Voltage	Traveling wave	2 Buses	No	No	No	230	No	[23]
[145]	2013	Current	Traveling wave	2 Buses	No	No	No	231	No	[21]
[146]	2022	Voltage	Traveling wave	2 Buses	No	Yes	Yes	230	No	[24]
[147]	2014	Voltage	Traveling wave	NM	No	No	No	230	Yes	NM
[148]	2013	Current	Wavelet & ANN	90 Buses	No	Yes	No	13.8	No	[12]
[149]	2018	Both**	Wavelet & ANN	IEEE 13 & IEEE 34	Yes	Yes	No	4.16 - 24.9*	No	[11]
[150]	2020	Current	ANN	IEEE 123	Yes	Yes	No	4.16	No	[24]
[151]	2015	Both**	ANN	4 Buses	No	No	No	400	No	[24]
[152]	2011	Both**	Wavelet & WANN	2 Buses	No	No	Yes	400	No	[24]
[153]	2006	Both**	Phasor & ANN	14 Buses	No	Yes	No	13.8	No	[11]
[154]	2018	Both**	Parameter & ANN	14 Buses	Yes	Yes	No	13.8	No	[12]
[155]	2005	Both**	Parameter & ANN	2 Buses	No	No	No	240 - 500*	Yes	[24]
[156]	2011	Both**	ANN	90 Buses	No	Yes	No	13.8	No	[12]
[157]	2022	Both**	XGBoost	IEEE-14	No	No	Yes	132	No	[24]
[158]	2018	Both**	ANFIS	24 Buses	No	No	No	25	No	[158]
[159]	2016	Both**	Fuzzy	227 buses	No	No	No	13.8	No	[24]
[160]	2020	Current	GPR	6 Buses	No	No	No	22	No	[24]
[161]	2016	Both**	SVM	2 Buses	No	No	Yes	400	No	NS
[162]	2014	Voltage	SVM & Traveling wave	2 Buses	No	No	Yes	230	No	[21]
[163]	2014	Both**	Zero Sequence Power	NM	No	No	No	22	Yes	NM

*Voltage level for a field test or second modeled network

**Both means that both voltage and current signals are used.

NS is for Not Stated

NM is for Not Modeled. These studies were verified with real data.

2.3.2.1 The popularity of HIF localization techniques

When the studies related to HIF location in literature are considered, 25 studies for finding fault sections and 40 studies for estimating the fault distance were found. Studies have been conducted since 2000 when HIFs started to be maturated. Although

fewer fault localization papers were published on the topic comparing the other subjects of HIFs, some outcomes can still be inferred.

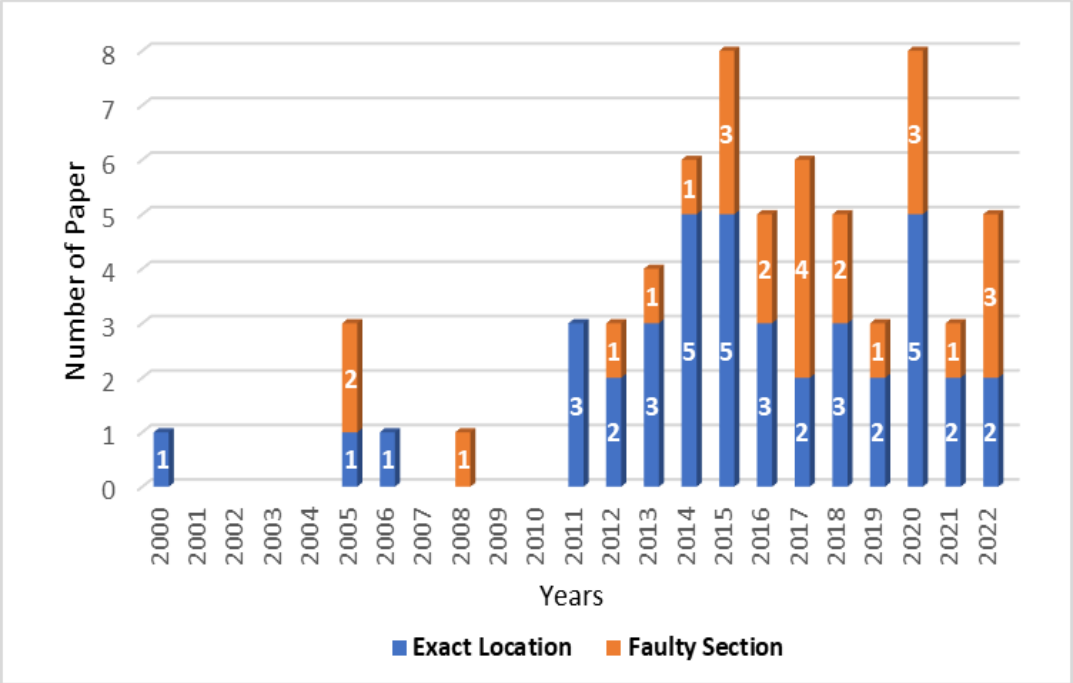


Figure 2.7 : Number of published papers by year.

As a first outcome, it can be declared that there has been a noticeable increase in interest in this field, particularly over the last decade, as shown in Table 2.2 and Table 2.3. The number of published papers in the literature continues to grow yearly. Figure 2.7 illustrates the yearly quantitative distribution of reviewed articles for each type of localization. As seen from the figure, only six papers were published between 2000 and 2010, while 59 papers were published in the following years.

2.3.2.2 Discussion of proposed HIF localization techniques in the literature

Various HIF localization methods were compared by examining several factors, including the applied methods, test systems, operating voltages, loading conditions, noises, and HIF models. These factors are summarized in Table 2.2 and Table 2.3, categorized by localization methods, which include identifying faulty sections and pinpointing exact locations.

When the literature is evaluated, it is observed that methods used in two different HIF location categories are varied. As discussed previously, two main approaches are considered: finding the estimated area and pinpointing the exact location. The pie

charts showing the distribution of proposed methods from the investigated papers are presented in Figure 2.8

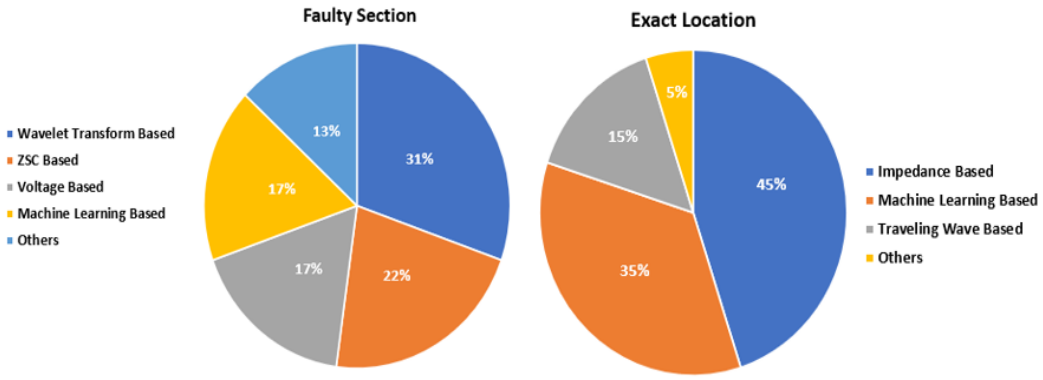


Figure 2.8 : Proportional percentages of methods used to find a faulty section and exact location by considering HIFs.

Although WT-based techniques are mainly used to estimate faulty zones, they do not necessarily offer a significant advantage over other methods. While the WT-based method was proposed in the mid of the 2010s, many researchers have preferred ZSC-based methods in recent years due to their straightforward implementation with relays. On the other hand, impedance-based and machine-learning methods are preferred in 82.5% of the studies for determining the exact location of HIFs. As mentioned previously, it is believed that the main reason for choosing impedance-based methods seems to be their ease of implementation despite their higher error rates compared to other methods. Machine learning has become the second most popular method due to its relatively uncomplicated applications and high success rates. However, 13 out of 14 reviewed machine learning-based studies did not test their methods with multiple networks to escape retraining the algorithm.

Similarly, measured signals in the two different categories of HIF location studies differ from each other. While both current and voltage signals are utilized in most of the pinpointing HIF studies, one of the power signals is generally preferred in faulty section localization studies.

Papers are also evaluated regarding their voltage levels as distribution and transmission network voltages. An analysis of Table 2.2 shows that over 90% of studies focused on distribution networks when aiming to find zones of HIFs. On the other hand, studies conducted for transmission networks increased from 10 to %50 in exact fault location studies, as indicated in Table 2.3. It is believed that this difference occurred due to the varying sizes of network configurations between distribution and

transmission networks. While distribution networks have lateral branches that create multiple pathways for faults, transmission networks are generally modeled as one line between two buses in simulation studies. Thus, identifying faulty sections in small networks as transmission networks will be a detection rather than a localization study. Based on this fact, it can be asserted that studies aimed at identifying faulty sections have focused on larger networks compared to pinpointing exact fault location studies.

Furthermore, it can be observed that most of the transmission networks (70%) used in precise fault location studies are modeled as two busses networks. Moreover, most studies did not test their methods in large and complex configurations. Even the studies use distribution networks, chosen simple networks. Only five out of 20 papers (25%) considered distribution networks and tested their methods in networks with more than 35 buses. This data supports the idea that authors choose simpler networks to avoid the challenges associated with accurately pinpointing HIF.

When validations of proposed methods are investigated, 58 out of the 65 papers presented pure simulation-based studies that did not test their methods with actual field data. 5 out of 7 remaining articles tested their methods with actual HIF signals in the simulations. The outcomes of these papers indicate that the error percentage in real data-based verifications is higher than in simulation-based studies. Therefore, it can be inferred that tests, including signals from stage faults, are necessary to ensure the performance of the proposed methods. Besides the actual data verification, some unavoidable conditions in the field were not accounted for in the studies. For instance, loading variation, an essential parameter particularly for the distribution networks, was not considered in all studies. Even though different loading conditions can affect the accuracy of almost all methods, only 23 papers investigated the influence of various loading conditions in proposed methods. Similarly, noise is also not taken into consideration for the majority (72.3%) of the studies. Even though it is inevitable in the actual power systems when measuring the voltage and current signals, just 12 from exact localization-based reviewed papers and six from faulty section-based reviewed papers checked the influence of noise to ensure the performance of a proposed algorithm. However, as an improvement, it can be stated that studies have placed greater emphasis on the effects of noise in recent years, according to Table 2.2 and Table 2.3.

Additionally, arcing high impedance faults are not taken into consideration in all studies to simplify it. Almost half of the studies for faulty section and exact location categories used the constant resistance model [24] as a fault rather than models representing HIF characteristics. In some reviewed studies, the error percentage increases with increased fault resistance even though they applied constant resistance as a fault. Thus, this percentage is expected to be higher with more complicated HIF models. Since the proposed methods in half of the studies do not consider realistic HIF models, their effectiveness cannot be guaranteed.

3. EXPERIMENTAL SETUP AND EVALUATION OF CAPTURED DATA

This section explains the test setup and recorded signals for each HIF type. Experiments for different kinds of HIFs are conducted in the Istanbul Technical University Fuat Külünk High Voltage Laboratory. Then, captured current signals are used to extract the characteristics of HIFs. The extracted characteristics from these captured HIF current waveforms were also investigated and visually proven with known behaviors.

3.1 Experimental Set-up

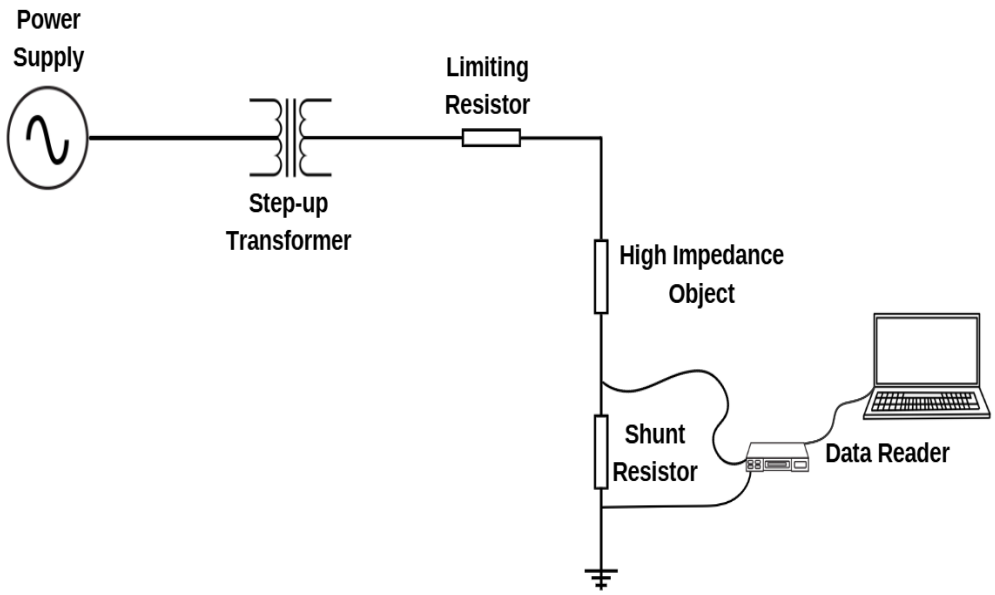


Figure 3.1 : Experimental test schematic of HIF.

In order to generate HIF signals, an experimental setup was created in a laboratory, and experiments were conducted. The test schematic for a high impedance fault is given in Figure 3.1 below. In this figure, a high voltage (HV) transformer with a power rating of 25 kVA is used to increase the test voltage to the distribution voltage level from the grid voltage. The voltage was applied up to 12 kV during the tests. A 25 kΩ

current limiting resistor is employed in series with high impedance to protect the transformer. Finally, a shunt resistor with a value of 25 ohms is used in the test system to record the current signals. Current signals were measured and recorded with a 5 kHz sampling frequency via Labview software for each created high impedance fault. These real HIF signals will then be used as fault data in the scenarios to be created in the selected branched distribution system in the simulation environment.

3.2 Tree Touching

The test setup images for the HIF event, which takes place when a tree or branch comes into contact with an overhead line, are shown in Figure 3.2 and Figure 3.3. A freshly cut tree was used for the testing. The arc began to ignite at approximately 3 kV, and shortly afterward, burning was observed at the contact point of the tree branch. Several tests were conducted, and HIF signals were recorded for further evaluation and simulation tests. One of these recorded signals is displayed in Figure 3.4 below.

Figure 3.2 : The first test set-up picture for a tree-leaning overhead line in the laboratory environment.

Figure 3.3 : The second test set-up picture for a tree-leaning overhead line in the laboratory environment.

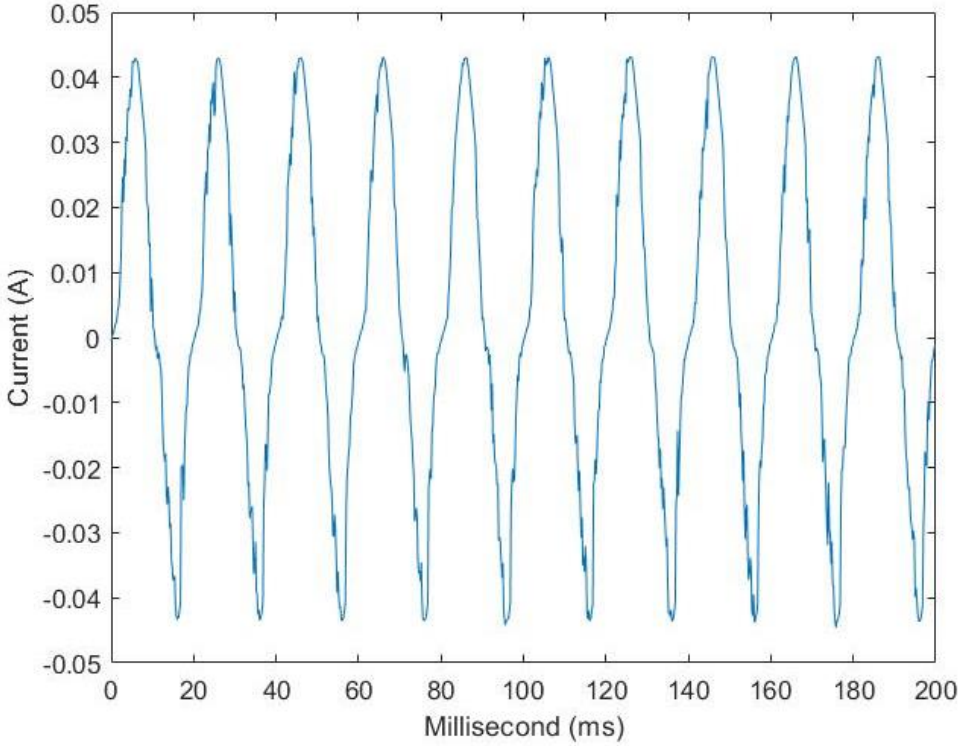


Figure 3.4 : Recorded HIF signal when a tree touches an overhead line.

To be sure about the test results, the captured signal for the tree branch was compared to one of the similar case signals from the literature. For both cases, two cycles of current signals are presented in Figure 3.5 and Figure 3.6. When one evaluates the signals, one notices the similarities in the distortions on zero crossing and peak values. Moreover, zero-crossing lengths are also almost identical.

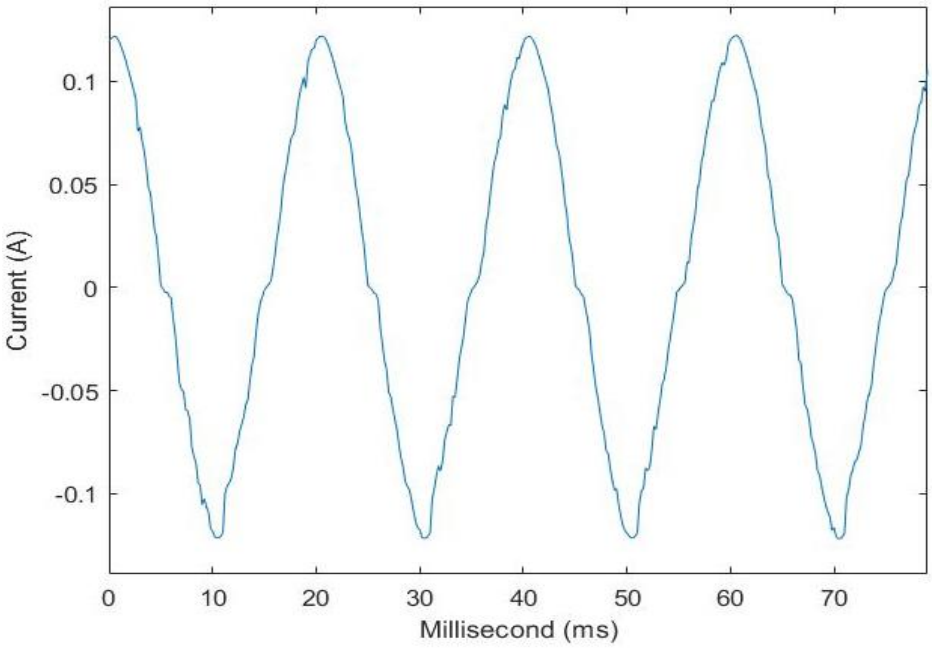


Figure 3.5 : Imported HIF signal when tree touches the overhead line.

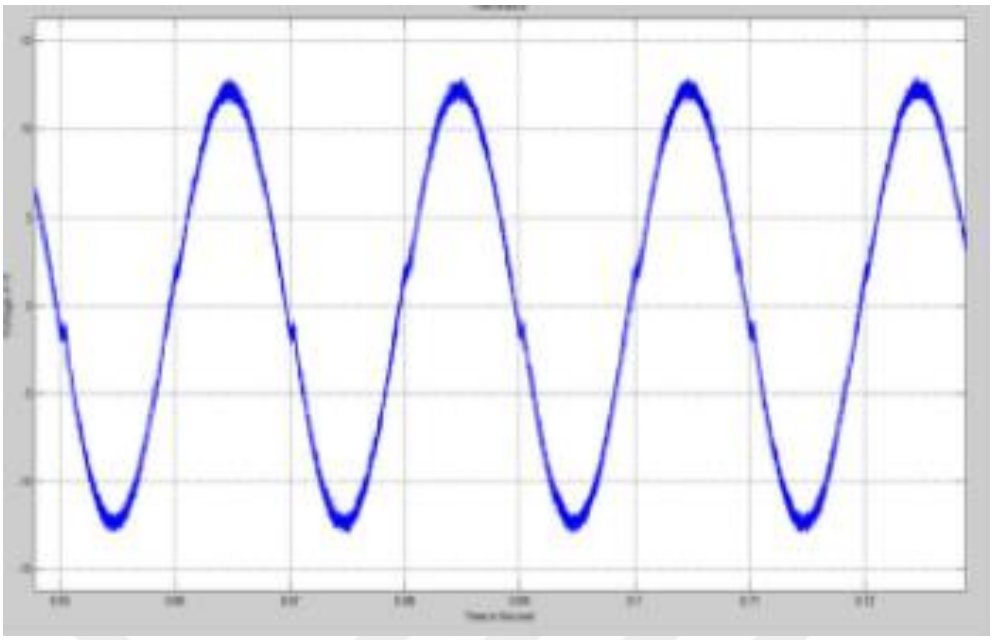


Figure 3.6 : HIF signal from literature when the tree touches the overhead line.

3.3 Broken Conductor

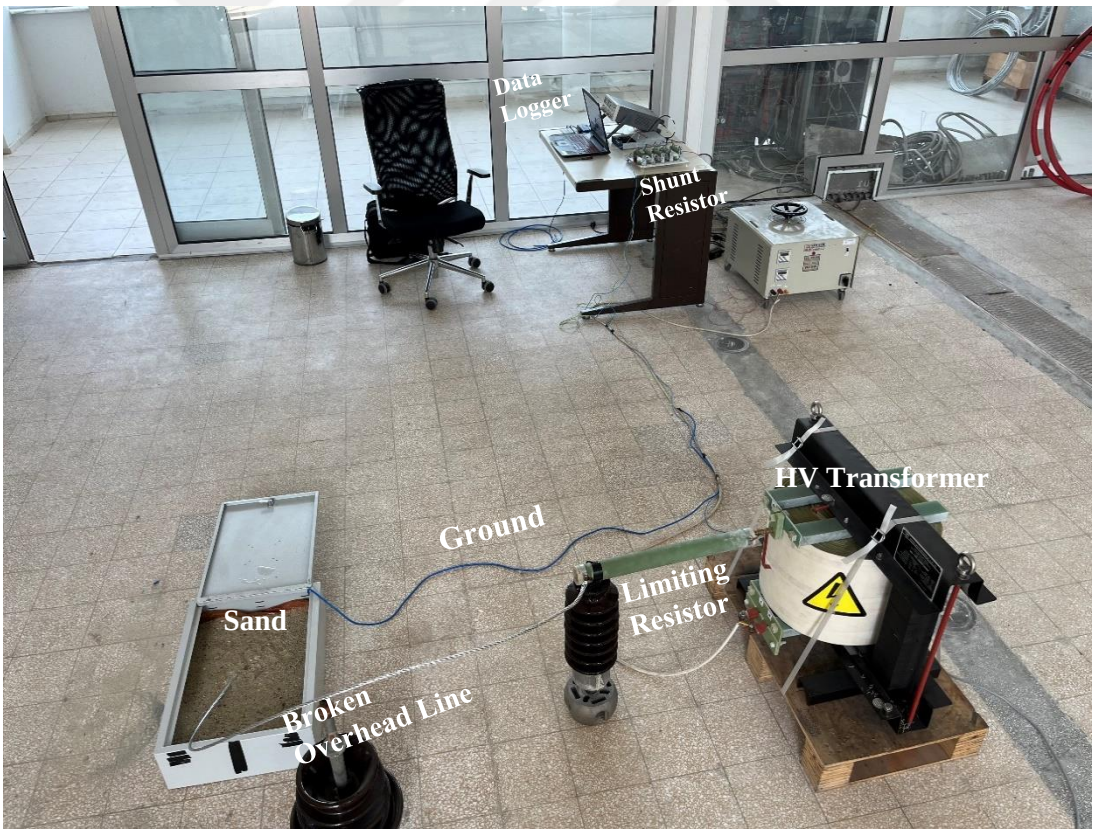


Figure 3.7 : HIF test picture when the contact surface is sand.

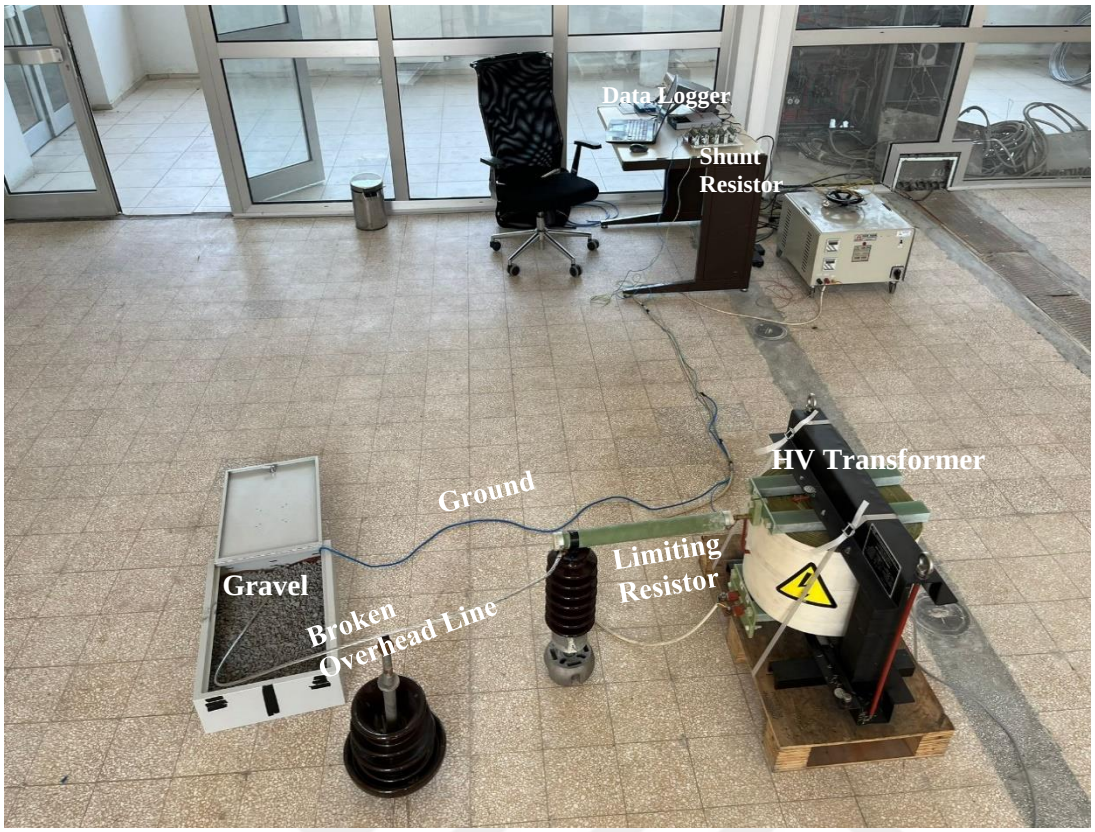


Figure 3.8 : HIF test picture when the contact surface is gravel.

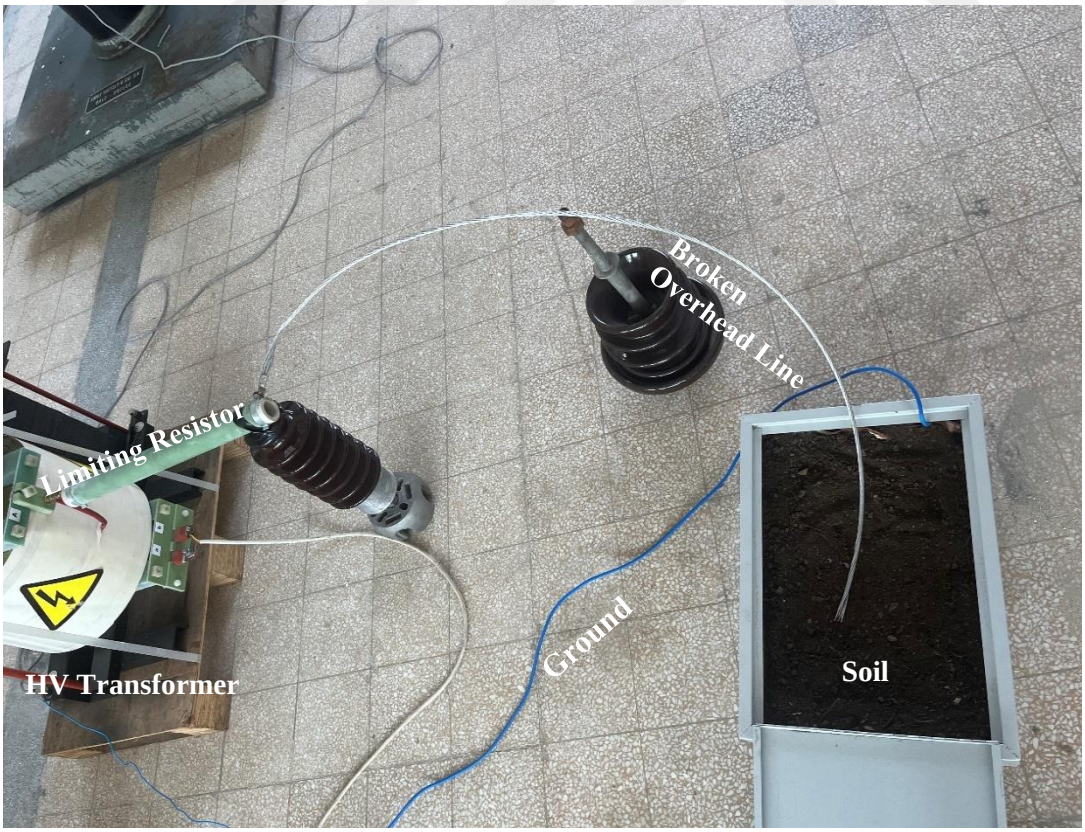


Figure 3.9 : HIF test picture when the contact surface is soil.

The applied voltage was up to 14 kV, and signals were recorded with a 5 kHz sampling rate using Labview software for broken conductor cases. Several scenarios involving changing surfaces were tested for a broken conductor. Sand, gravel, and soil were selected as the surfaces for testing. The test setup pictures for a fault, when the overhead line conductor breaks and touches the surface of sand, gravel, and soil, are presented in Figure 3.7, Figure 3.8, and Figure 3.9 respectively. While the depth of the contact surface was 5 centimeters (cm) in most of the tests, 2.5 cm and 7.5 cm depths were also tested to assess their impact.

3.3.1 Gravel

Initial experiments were conducted on dry gravel using various voltage levels. The voltage signals were captured through a 25 Ω shunt resistor for different voltage magnitudes and depths, as illustrated in Figure 3.10 and Figure 3.11. The data indicate a consistent increase in current values relative to the applied voltages.

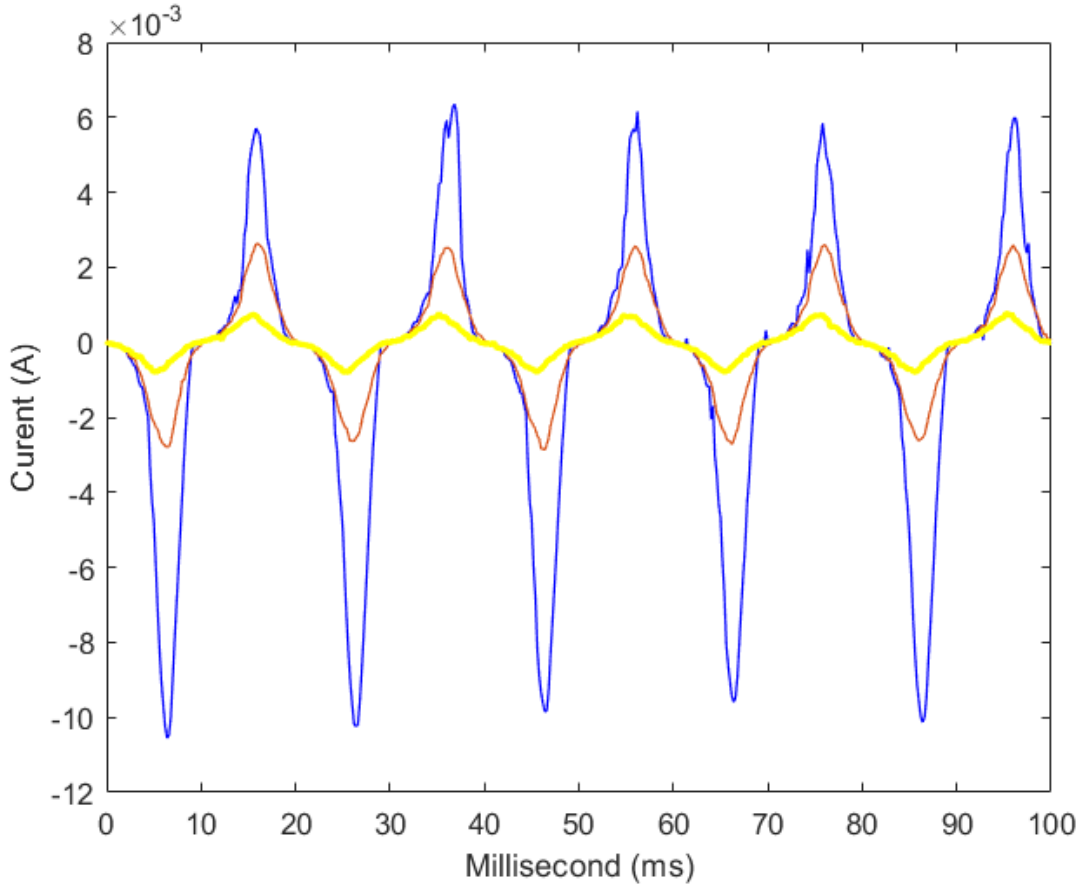


Figure 3.10 : Recorded HIF signals for different applied voltages when the surface is 5 cm dry gravel (Blue: 11.5 kV, Red: 7 kV, Yellow: 4 kV).

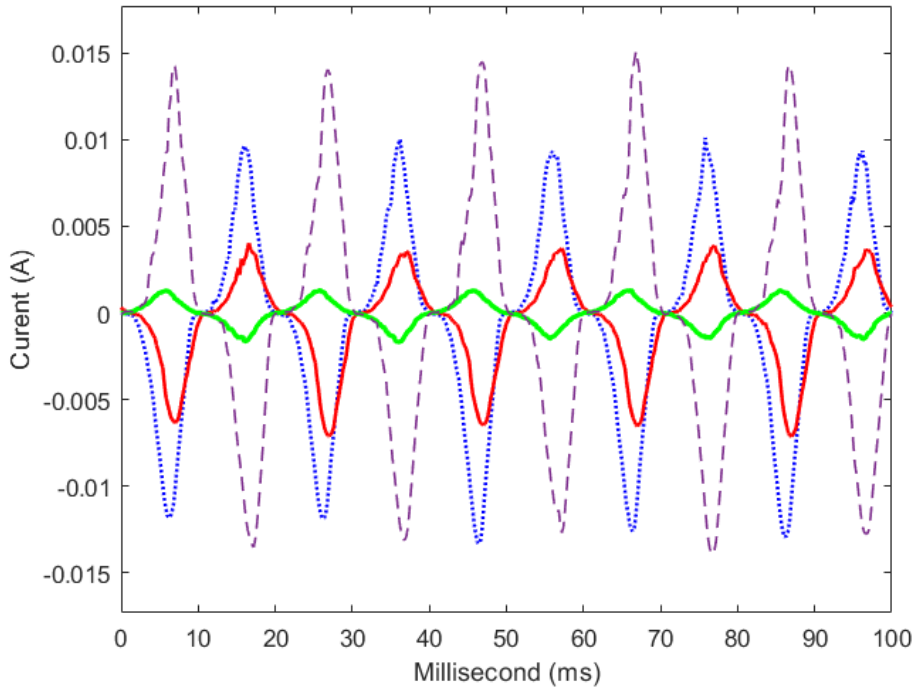


Figure 3.11 : Recorded HIF signals for different applied voltages when the surface is 7.5 cm dry gravel (Purple:11.5 kV, Blue: 9.5 kV, Red: 7 kV, Green: 4 kV).

Furthermore, an FFT was performed to analyze all recorded signals, revealing the presence of the third harmonic within the fault signals. The FFT results for two distinct voltage levels are presented in Figure 3.12. It has been observed that the 3rd harmonics are dominant in the fault current along with the 5th harmonics, although the percentage is different at various voltages.

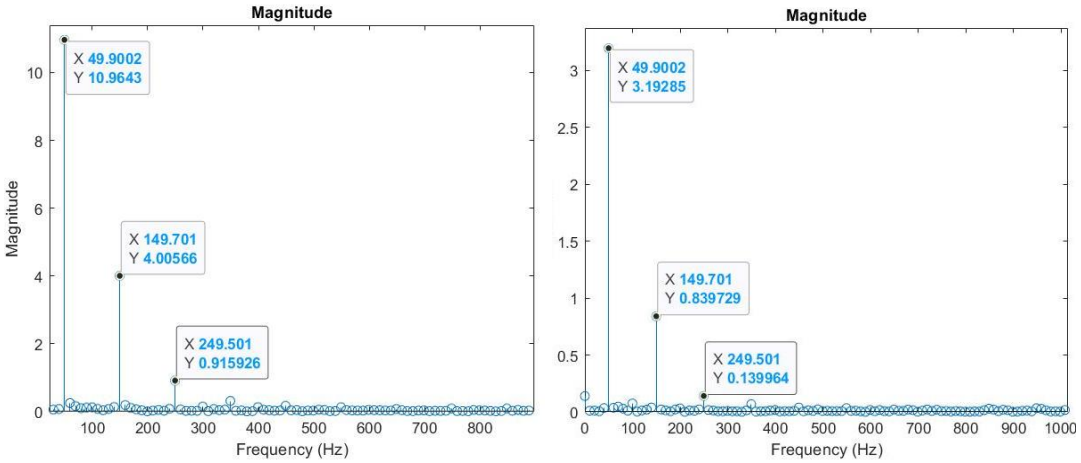


Figure 3.12 : FFT results for recorded HIF signals when the surface is dry gravel (Left: 7 kV, Right: 4 kV).

In addition to the investigations conducted on dry gravel for different voltage levels, HIF signals obtained at several depths were compared using a voltage of 4 kV, as

shown in Figure 3.13. As the depth decreases for the same applied voltages, higher currents are observed, as expected.

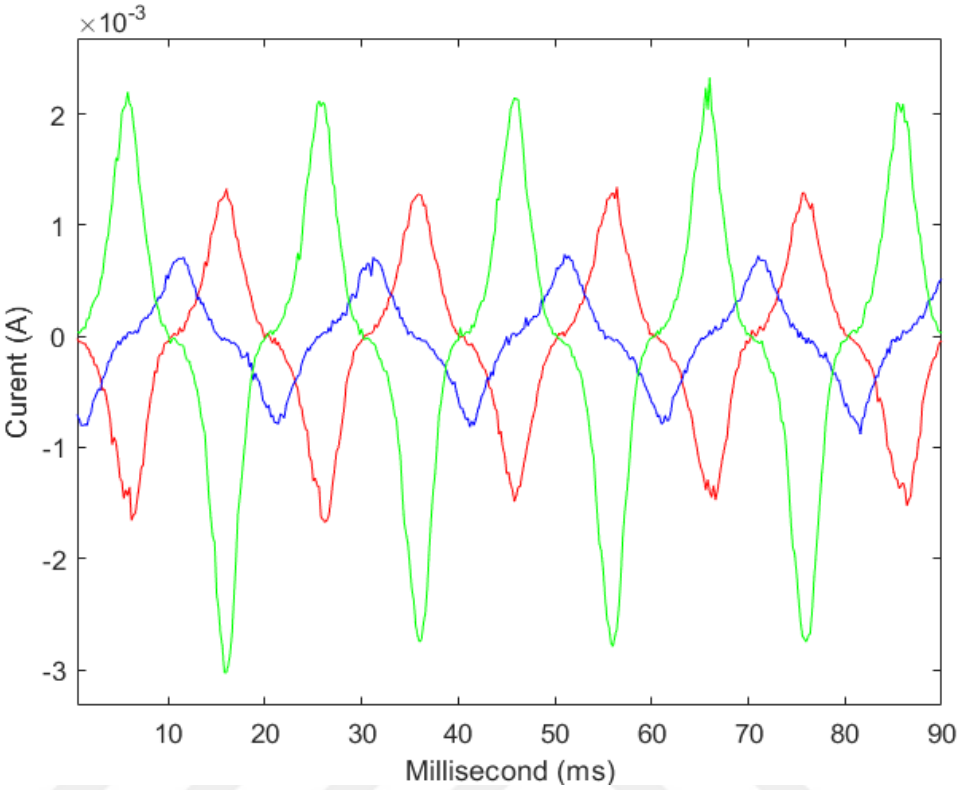


Figure 3.13 : Recorded HIF signals for different depths (2.5-5-7.5 cm) when the surface is dry gravel.

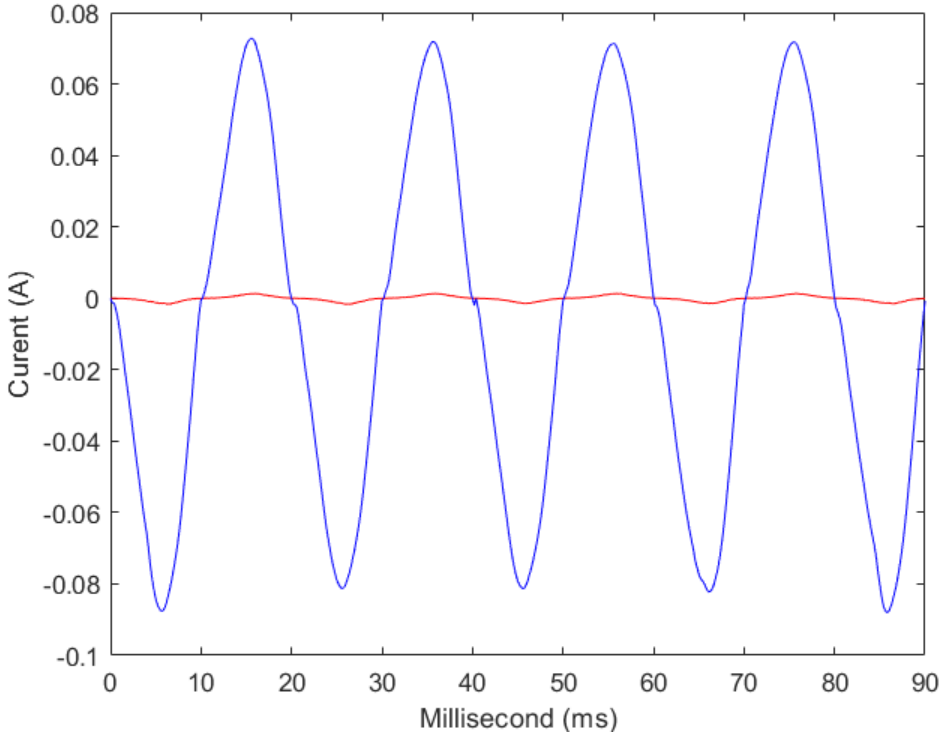


Figure 3.14 : Comparison of recorded HIF signals for when the surface is dry and wet gravel.

Subsequent tests were carried out with wet gravel at a depth of 7.5 cm. Due to the rapid increase in voltage values, a maximum voltage of only 4 kV was applied for these tests. Figure 3.14 presents a comparison of the HIF waveforms from dry and wet gravel at a depth of 7.5 cm. Based on the results, higher current flows were observed on the wet surface.

3.3.2 Sand

Aside from gravel, laboratory tests were conducted on sand as well. The measurement of soil conductivity was not conducted; however, as indicated in the literature, it is typically between 1 and 100 mS/m. The sand employed in this thesis is derived from a coastal beach environment. The depth measured 5 cm, and various voltages were applied when the contact surface was dry sand. The recorded fault signals are illustrated in Figure 3.15. When comparing these signals to those from gravel, the signals from sand displayed similar characteristics but at higher magnitudes.

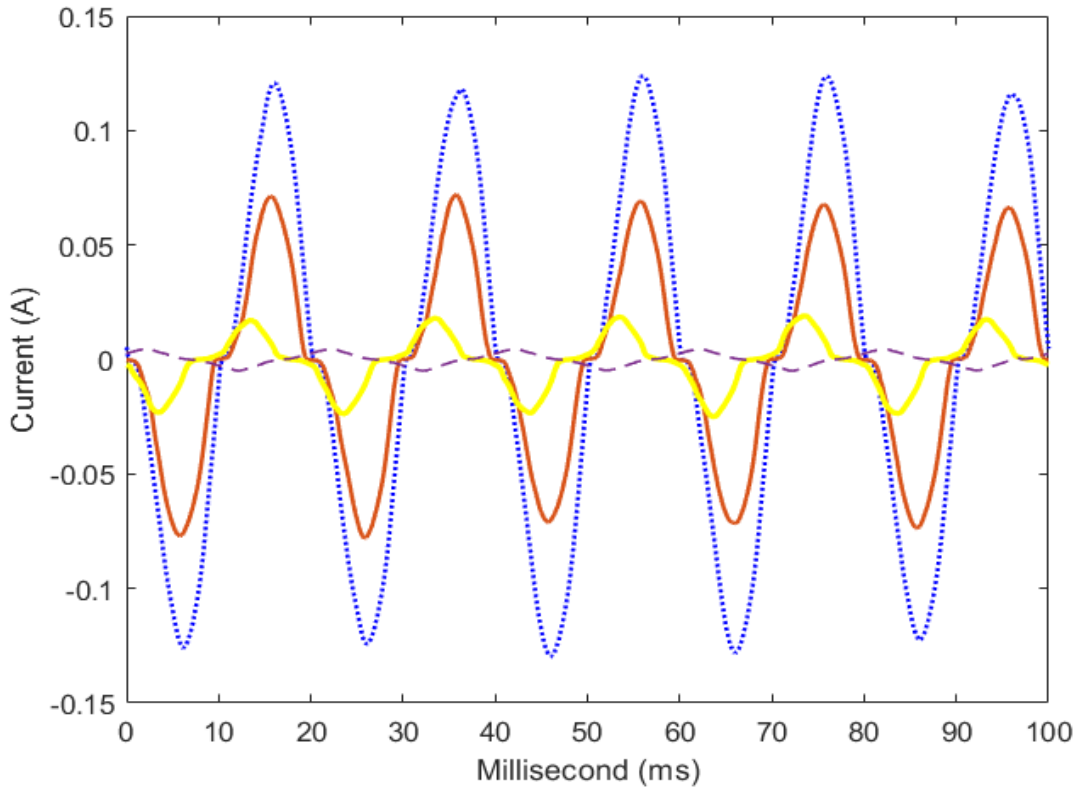


Figure 3.15 : Recorded HIF signals for different applied voltages when the surface is 5 cm dry sand (Blue:11.5 kV, Red: 9.5 kV, Yellow: 7 kV, Purple: 4 kV).

Moreover, wet sand was tested in 4kV applied voltage. As expected, the recorded current signals exhibited higher magnitudes. For better understanding, a comparison

was made between dry sand at 11.5 kV and wet sand at 4 kV. An illustration of both signals is shown in Figure 3.16.

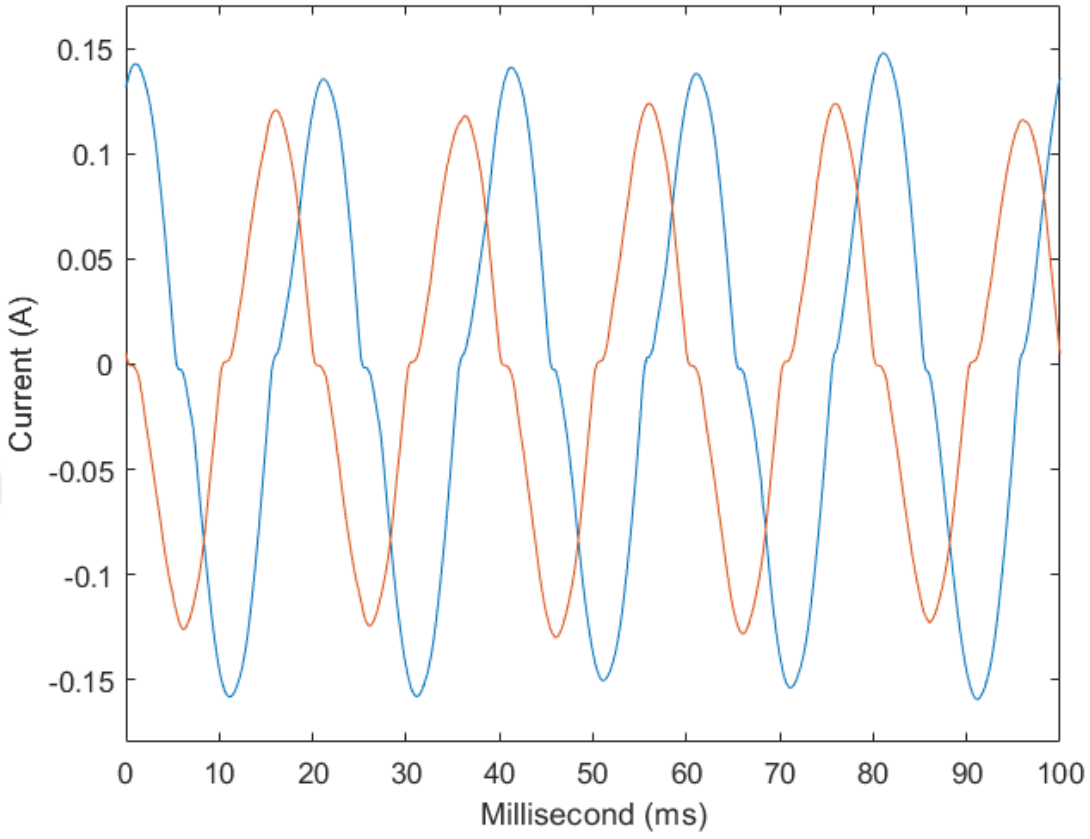


Figure 3.16 : Comparison of recorded HIF signals for when the surface is dry sand (Applied voltage: 11.5 kV, Color: Red) and Wet Sand (Applied voltage: 4 kV, Color: Blue).

3.3.3 Soil

Soil was selected as the final test surface. Its lower resistivity than other surfaces resulted in higher current magnitudes in the fault signals. Therefore, only 4 kV and 7 kV were applied, and the recorded signals are illustrated in Figure 3.17. Additionally, the frequency spectrum of the recorded signal at 4 kV is provided in Figure 3.18. Consistent with other surfaces, 3rd harmonic values appear to be dominant.

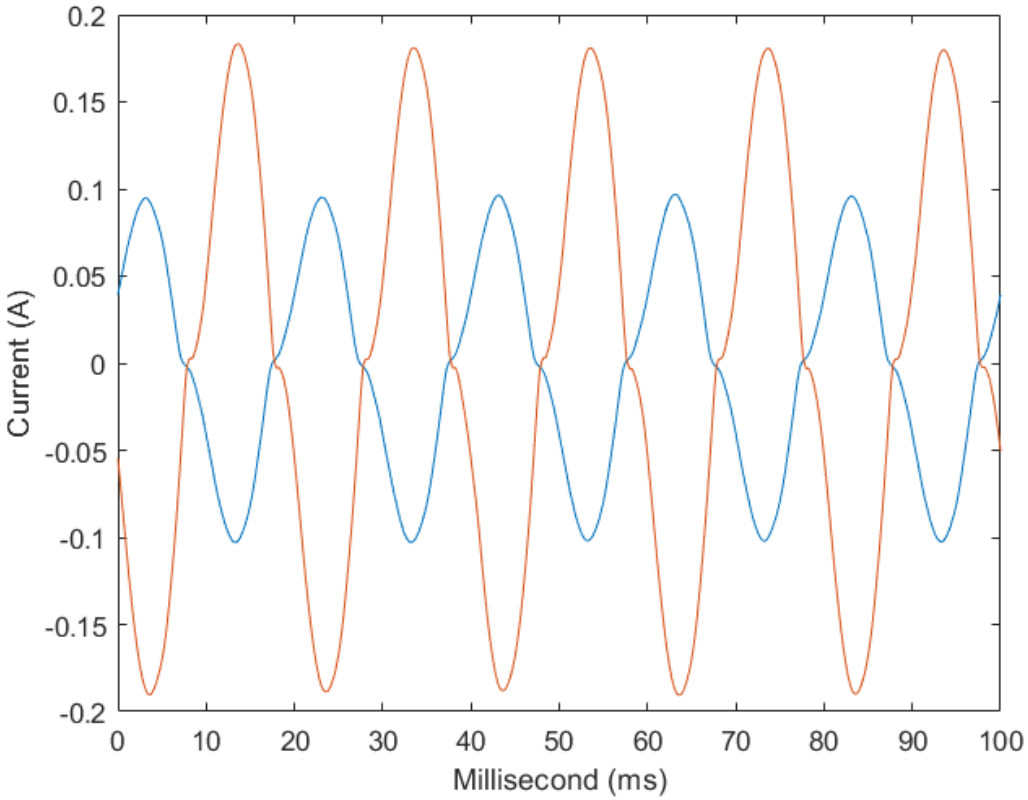


Figure 3.17 : Recorded HIF signals for different applied voltages when the surface is 5 cm soil (Red: 7 kV, Blue: 4 kV).

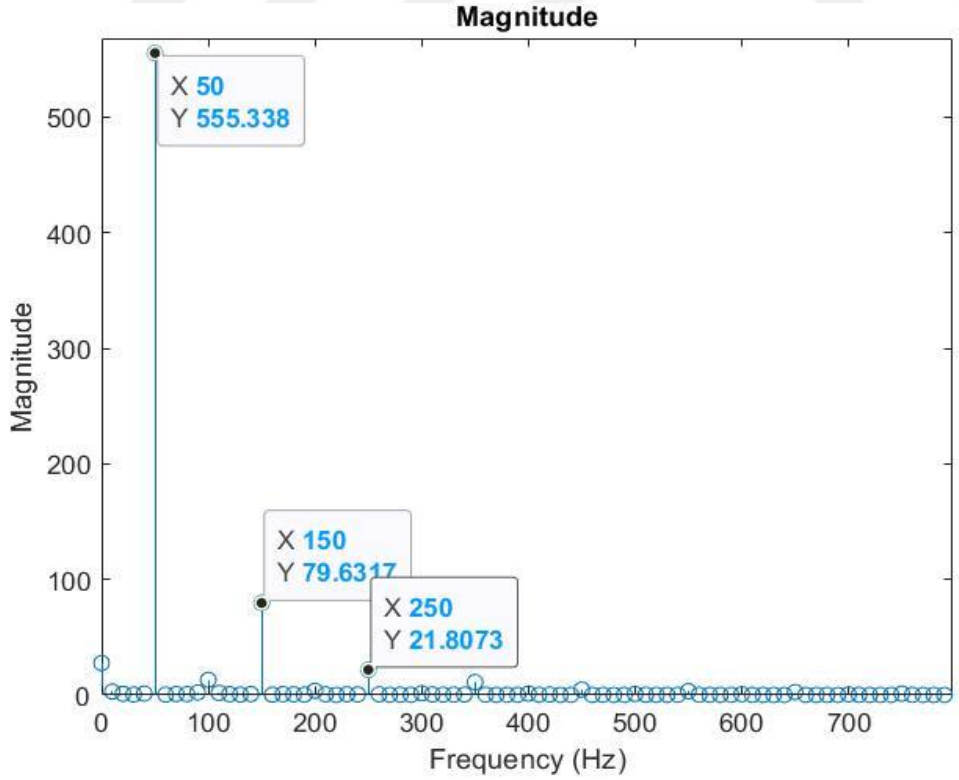


Figure 3.18 : FFT results for recorded HIF signals when the surface is soil (Applied voltage: 4 kV).

3.4 Characteristics of Captured Signals

As mentioned in the literature, HIFs have several characteristics. This section presents these specific features of HIFs by evaluating the captured current signal from laboratory tests.

3.4.1 Low fault current

The amplitude of the measured fault current is low due to the high impedance of the fault path caused by tree and surface contacts. This characteristic is one of the most fundamental features of HIF and one of the main reasons why it's challenging to detect such faults with traditional protection devices.

3.4.2 Asymmetry

The positive and negative half-cycles of HIFs display asymmetrical behavior. This characteristic of asymmetry, seen in recorded signals, is presented in Figure 3.19.

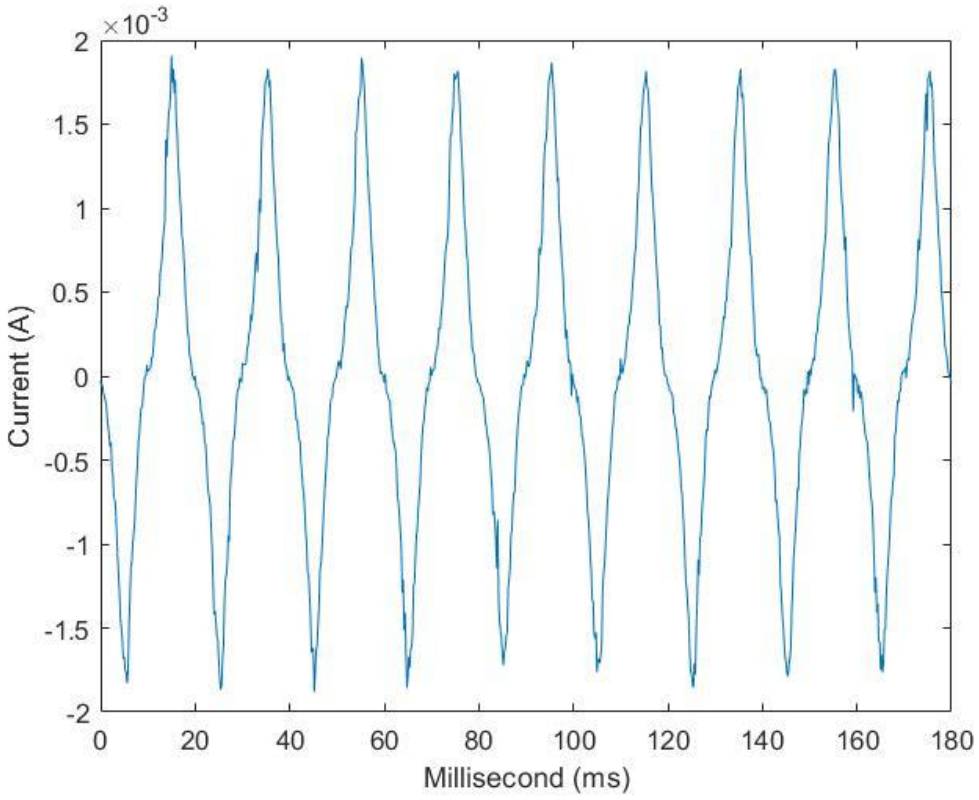


Figure 3.19 : Asymmetry behavior of HIF signal.

3.4.3 Nonlinearity

Another phenomenon to found is the nonlinear characteristic occurring around the zero crossing, which leads to the creation of harmonics in the HIF current. Distortions at the signal are apparent, especially in zero-crossing, as shown in Figure 3.20. These distortions and non-linearity characteristics create harmonics in the faults.

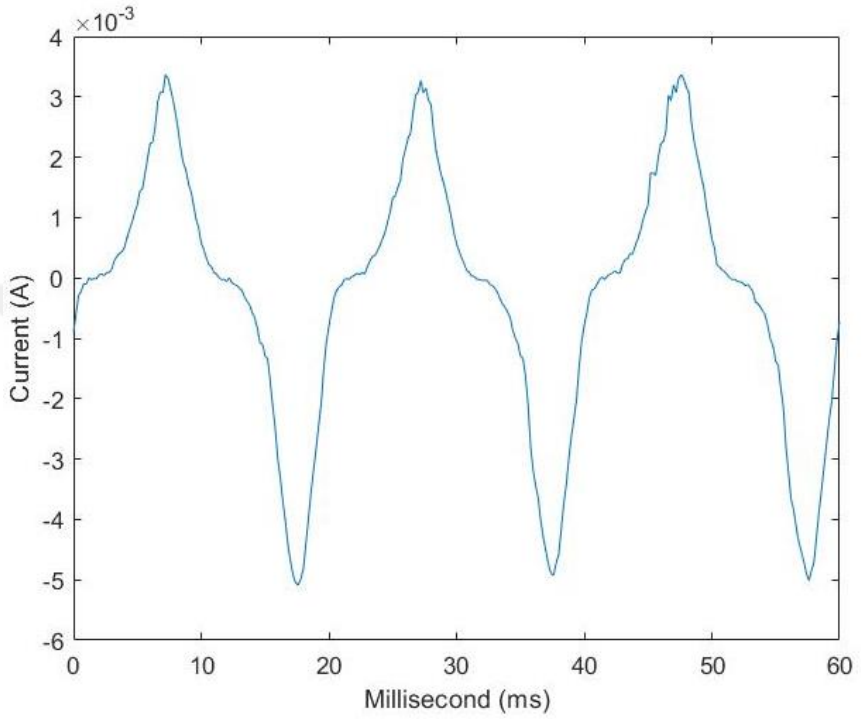


Figure 3.20 : Nonlinear behavior of HIF signal.

3.4.4 Build up and shoulder

One of the documented features of HIF is build-up and shoulder. In this feature, the current increases several cycles and remains steady for a few cycles. The same characteristic is observed in recorded signals. One of the signals showing this feature is presented below in Figure 3.21.

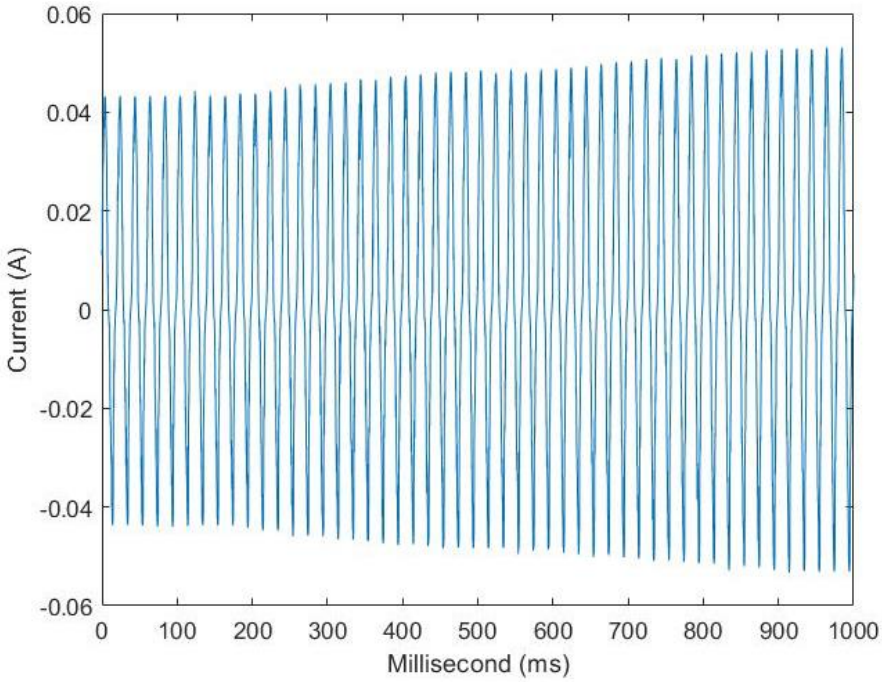


Figure 3.21 : Build-up and shoulder behavior of HIF signal.

3.4.5 Randomness

It's essential to note the random nature of the HIF current. The impedance of the faulted path randomly changes, leading to variations in current amplitude, as illustrated in Figure 3.22. Here, impermanent physical contact of the line with high-impedance objects and the arcing creates an unpredictable fluctuation in the current signal.

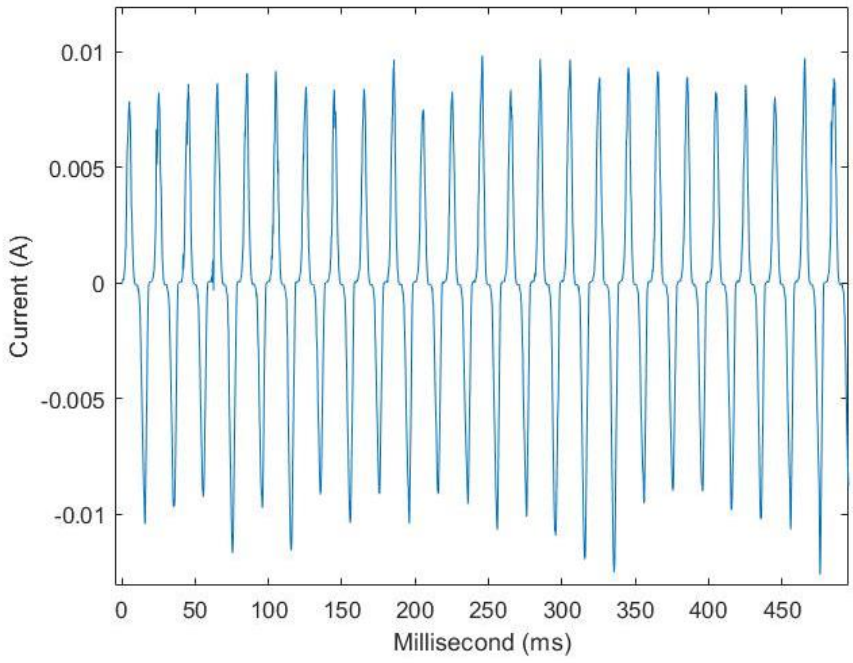


Figure 3.22 : Randomness behavior of HIF signal.

3.4.6 Intermittence

The HIF current is not consistently stable because the arc goes through periods of extinction and reignition, resulting in disturbances in the fault current. Additionally, the fault impedance fluctuates due to chemical reactions and other factors. These factors result in the intermittence phenomenon, as shown in Figure 3.23.

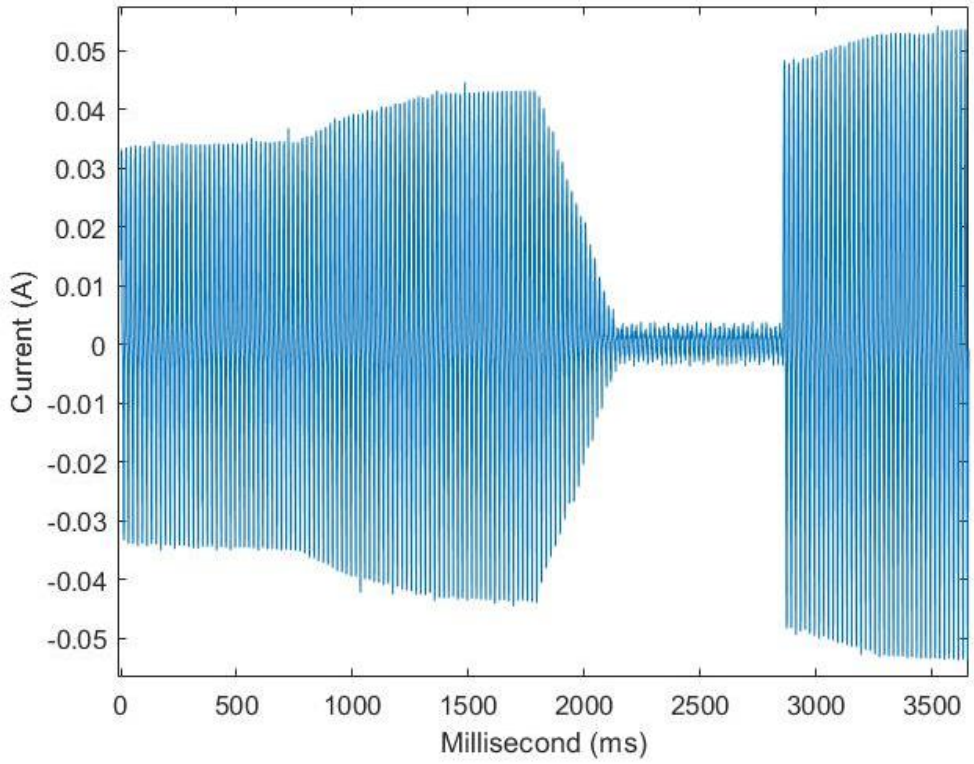


Figure 3.23 : Intermittence behavior of HIF signal.

4. HIF DETECTION STUDIES WITH MACHINE LEARNING METHODS IN BRANCHED DISTRIBUTION NETWORK

High impedance faults are complex faults that affect the stability and reliability of the distribution networks. Based on reviewed studies, machine learning methods have been used extensively to detect such faults due to their superiority to others in the detection performance. However, their performances were not evaluated thoroughly in the studies. Most published papers didn't consider actual fault signals and all possible cases that can occur in the distribution networks. Therefore, events having similar characteristics with HIFs were also considered in the simulation.

Besides different cases, two scenarios were also tested with various machine learning algorithms, such as ANN and SVM. The initial scenario involved testing two input features from the literature in machine learning algorithms with three different datasets. These datasets included HIFs created based on predefined models, actual recorded signals from laboratory tests, and both. Since chosen input features from the literature utilized wavelet coefficients, two mother wavelets were tested to understand their effects on the accuracy. Moreover, different configurations of machine learning methods were evaluated.

As a second investigation, the influence of the noise on pre-trained algorithms was tested. This sensitivity analysis considered input parameters from the previous scenario. Since real-world data often contains noise, 20 dB, 40 dB, and 60 dB signal to noise ratio (SNR) noises were added to signals to analyze the performances of machine learning methods.

All the simulations were conducted using the IEEE 34-bus test benchmark radial distribution network, modeled in a MATLAB/Simulink environment. This unbalanced test system is a real network located in the United States. It includes a mix of single-phase and three-phase loads, overhead lines, voltage regulators, and capacitor banks. Therefore, it can be considered a realistic and complex distribution network. Current

waveforms were recorded on the source side of the network (Bus 802). An illustration of the IEEE 34 bus test benchmark is presented below in Figure 4.1.

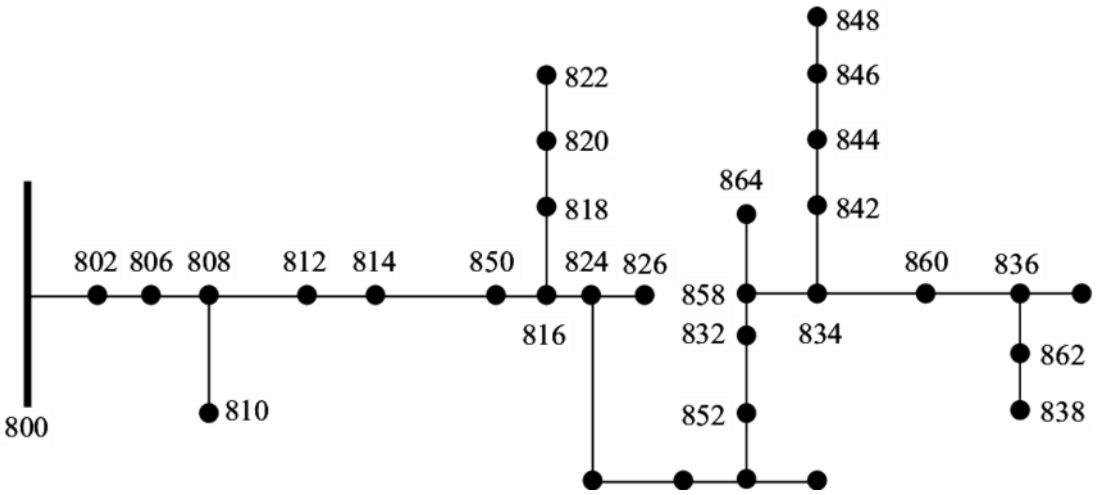


Figure 4.1 : IEEE 34 bus test benchmark.

In the following, tested machine learning algorithms, discrete wavelet transforms, and created various test cases are explained. Then, the results of the different machine learning algorithms considering several scenarios are discussed in detail.

4.1 Theory Of Utilized Machine Learning

The main objective of machine learning is to intelligently analyze the data and address problems by creating efficient learning techniques. Machine learning encompasses a wide range of techniques, divided into categories such as supervised, unsupervised, semi-supervised, and reinforcement learning.

4.1.1 ANN

ANN is one of the most popular machine learning methods. The term "Artificial Neural Network" is derived from biological neural networks. ANN consists of interconnected neurons in different layers. The artificial neurons are connected through specific weights (ω) adjusted in the learning process [164]. Artificial neural networks consist of 3 layers, which are input, hidden, and output layers. Figure 4.2 provides a general diagram of ANN.

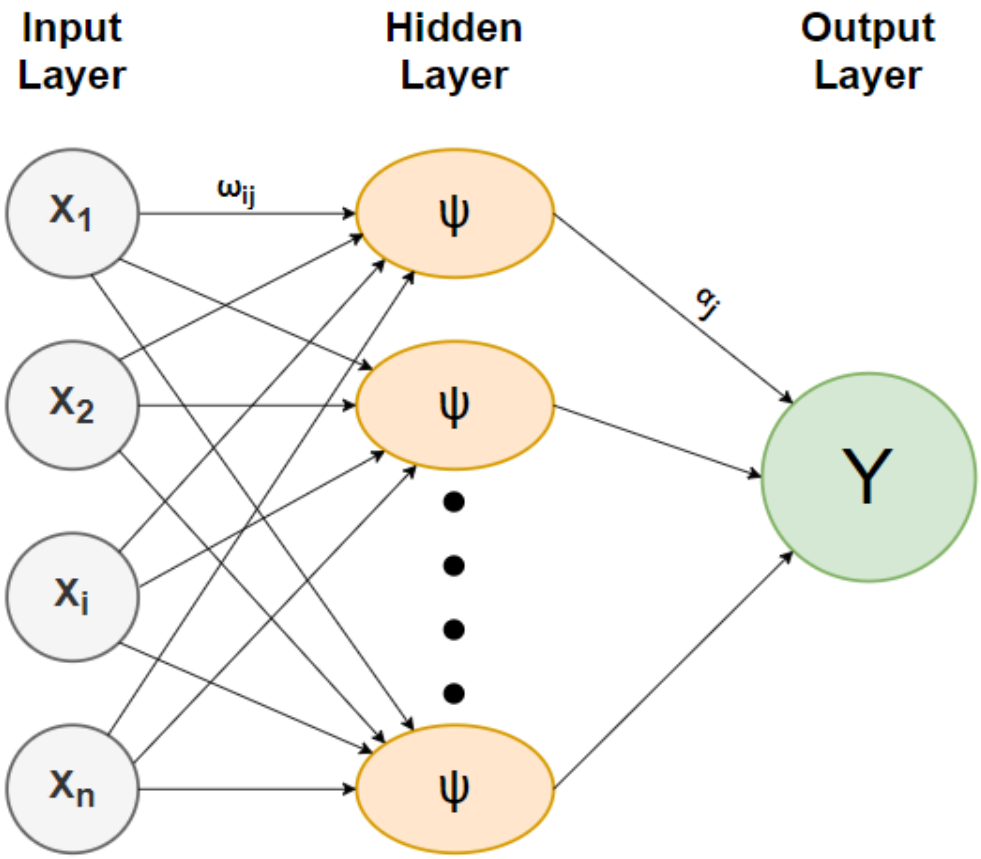


Figure 4.2 : Artificial Neural Network with 1 hidden layer.

Each input (X_i) passes through neurons in the hidden layer after multiplying with their respective weights (ω_{ij}). Next, these values go through activation functions (ψ) in the hidden layer, and the weighted (α_j) results are summed to determine the output value (Y) in the output layer. Throughout the training process, the network iteratively modifies its weights and biases. The process continues until the error in training data reaches an achievable minimum value. In other words, the process continues till the difference between the predicted output and the actual target is minimized. There are different types of ANNs. One of them is backpropagation. Backpropagation (BP) is widely recognized as one of the most effective methods in the ANN literature. Therefore, the backpropagation algorithm [165] is applied in this study for a supervised learning neural network. It has one hidden layer and 100 neurons.

4.1.2 SVM

Support Vector Machine is a well-known machine learning method used mainly in classification problems. The main working principle of the technique is creating a line to separate various classes of data sets. This line is called a hyperplane. The objective

is to find the best hyperplane that maximizes the margin between data sets of categories. Several types of kernel functions, such as polynomial, linear, and Radial Basis Functions (RBF), can be used in SVM to separate training data in different (higher) dimensional spaces [166]. A linear kernel function that gave better results for HIF detection is chosen in this paper, as shown in Figure 4.3.

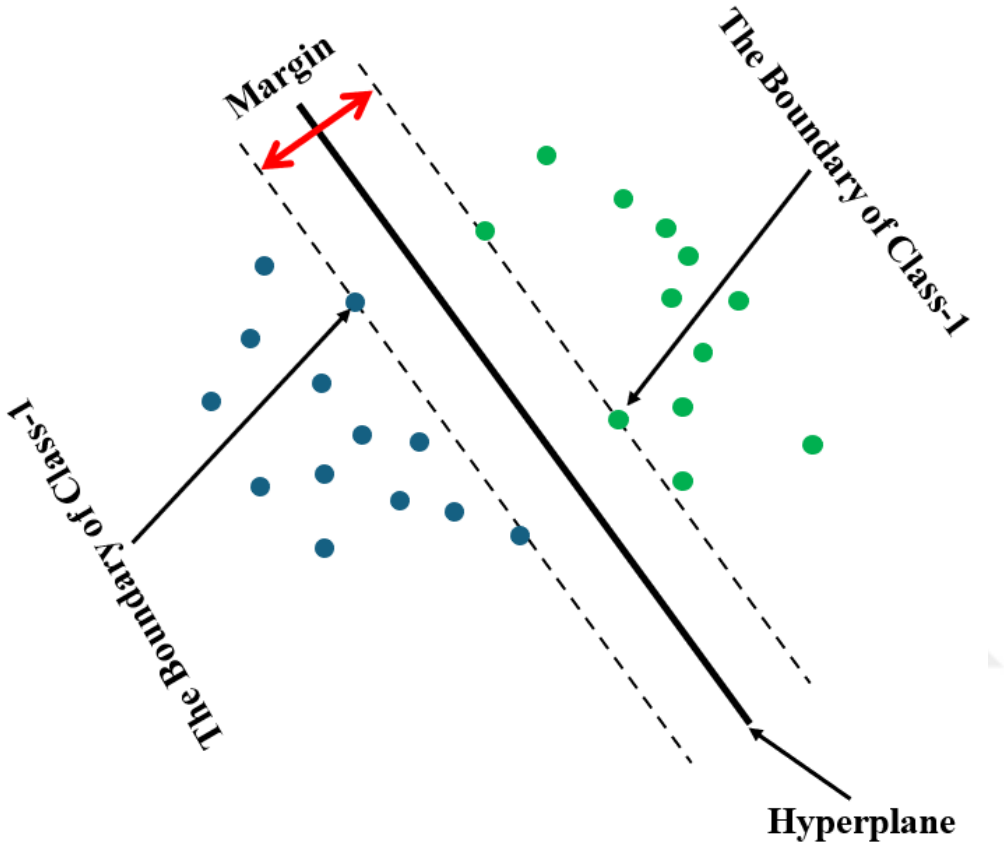


Figure 4.3 : Linear Support Vector Machine.

4.2 Theory Of Discrete Wavelet Transform

The discrete wavelet transform was created in 1988 [168]. Since then, this method has been widely used to address power network issues. DWT decomposes a signal into a series of dilated and translated versions of the mother wavelet function as formulated in (4.1).

$$DWT(\mathrm{f},\mathrm{m},\mathrm{n}) = \frac{1}{\sqrt{\mathrm{a}_0^{\mathrm{m}}}} \sum_{\mathrm{i}} \mathrm{X}_i * \varphi\left(\frac{\mathrm{n} - \mathrm{ib}_0\mathrm{a}_0^{\mathrm{m}}}{\mathrm{a}_0^{\mathrm{m}}}\right) \tag{4.1}$$

In the formula, the variable "m" is an integer that controls the scaling (a) and translation (b) parameters. X(k) represents the input signal, and " φ(n)" stands for the

mother wavelet. The Multiresolution Analysis is the most commonly utilized technique for carrying out the DWT. This approach breaks down the signal into components that represent both low and high frequencies, known as approximation (A) and detail (D) coefficients. Figure 4.4 illustrates the structure of a three-level DWT algorithm applied to an input signal.

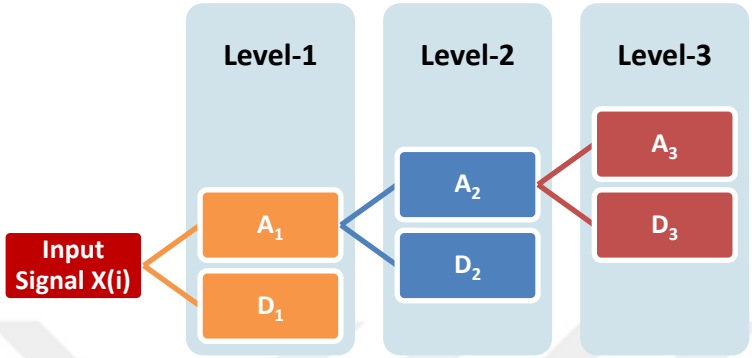


Figure 4.4 : Structure of three-level DWT **(**A: Approximation, D: Detail).

4.3 Simulations Cases

Since machine learning methods were selected for HIF detection, a database was created to train the algorithms. Thus, various non-HIF distribution power system events, such as shunt capacitor bank switching, linear load switching, rectifier load switching, and low impedance fault (short circuit), were simulated with HIF cases to investigate the robustness of algorithms. Details of these events are provided in the following subsections.

4.3.1 Actual HIFs

Recorded HIF signals that occurred due to tree-touching lines were imported into the simulation directly as actual HIFs. These signals were modeled as a current source. The magnitudes were scaled to vary between 5 %and 20% of the network's nominal current. Twenty different actual HIFs were created and added to the simulations.

4.3.2 Modeled HIFs

For the high impedance fault model, a simplified Emanual model is used [29]. Two unequal resistance R1 and R2 series with a diode and direct current (DC) voltage sources were implemented parallelly in this model as shown in Figure 4.5. Range of

R1 and R2 values were assigned such that the magnitude of fault current varies between %5 and %20 of the nominal current. Then, 20 different HIF cases were modeled and simulated.

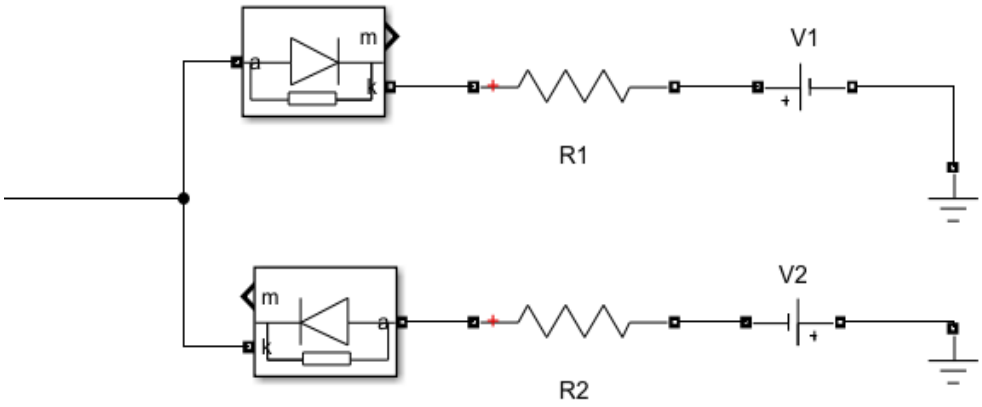


Figure 4.5 : Simplified Emanual HIF model.

4.3.3 Shunt capacitor bank switching

The switching of capacitor banks is one of the events in distribution networks that are inevitable. Furthermore, such events show similar characteristics to HIF signals. Therefore, they should be considered while testing the effectiveness of HIF detection methods. Hence the test system already has capacitor banks placed on buses 844 and 848, switches were inserted in series with these capacitors, and the simulation tests were conducted. Capacitor bank capacities for full and half loads were analyzed for each case, resulting in four combinations (2x2=4 cases).

4.3.4 Linear load switching

test the efficiency of HIF detection methods. New loads with switches were added to the system to simulate load switching. Both one-phase and three-phase loads between 0 % and 15 % of the nominal load in a step of 2.5 % were modeled (6 Cases). Moreover, the same loads were simulated as increasing loads started from their half-assigned value with the closing of switching to the total assigned magnitude of load during 150 ms (6 Cases).

4.3.5 Rectifier load switching

HIFs have nonlinearity characteristics that cause harmonics up to 600 Hz in current signals [13]. Therefore, nonlinear loads exhibit behaviors close to HIFs and should be studied in HIF detection studies to prove the performance of the proposed methods. In this study, a six-pulse rectifier with DC loads was modeled. Similar to the linear loads, both direct and increasing loads were simulated, and their values were changed from 0 % to 15 % of the nominal load in a step of 2.5 % (2x6=12 Cases).

4.3.6 Low impedance fault (short circuit)

Studies indicated that more than 80 % of faults in a distributed system are single-phase to-ground faults [169]-[170]. Thus, such events should be examined to test the robustness of the methods. In addition to high impedance faults, simulations of low impedance faults in one phase to the ground were carried out for 20 cases in different locations and used in the database to train algorithms.

4.3.7 Noise

There is another phenomenon that affects the verification of proposed detection methods. This inevitable factor is noise. In a real system, each recorded data comes with noises. To understand its impact on detection methods, noises of SNR 20dB, 40 dB, and SNR 60 dB were added to signals to be trained.

4.4 Simulations Results

All the created cases for training and testing data have been listed in Table 4.1. The duration of cases was 150ms, and the beginning time was changed in the cycle slot (16 instances in one cycle) for each simulated case. As indicated in Table 4.1, 1408 cases were simulated. Methods were tested both with and without considering actual HIF data. Therefore, different databases are created. The first database consists of only modeled HIF, and the second one involves only actual HIF signals. On the other hand, a third combines all HIF cases. After the creation of databases, 70% of cases were used for training the algorithms, while 30% were reserved for testing and validation.

Table 4.1 : Number of cases in terms of related events.

Simulated Cases	Number of cases
High Impedance Fault-Real	20 x 16 =320
High Impedance Fault-Modeled	20 x 16=320
Low Impedance Fault	20 x 16=320
Shunt capacitor bank switching	4 x 16=64
Linear load switching	12 x16=192
Rectifier load switching	12 x 16=192

4.4.1 Scenario 1: Results of published methods for different HIF datasets

Two machine learning methods with different configurations were assessed across three databases. Moreover, two distinct input approaches from published studies were selected for evaluation. The first approach involved using the standard deviation of the wavelet coefficients of seven levels. The second approach utilized the energy sum of the first four levels of discrete wavelet coefficients as input. The Db4 and sym5 wavelets were employed as the mother wavelets to derive the coefficient values. An initial investigation was conducted using an ANN with different activation functions and layer configurations while the epoch was set to 100. The results for the first approach are presented in Table 4.2.

Table 4.2 : ANN results when inputs are the standard deviation of coefficients for seven levels.

Configuration	Actual HIF	Modeled HIF	Both
Activation: Logistic Layer: (100,1) Mother Wavelet: db4	95.00	99.20	92.27
Activation: Logistic Layer: (50,2) Mother Wavelet: db4	62.75	81.69	74.88
Activation: Tanh Layer: (100,1) Mother Wavelet: db4	69.92	76.63	71.90
Activation: Tanh Layer: (50,2) Mother Wavelet: db4	91.12	96.12	85.82
Activation: Logistic Layer: (100,1) Mother Wavelet: sym5	94.29	99.63	95.15
Activation: Logistic Layer: (50,2) Mother Wavelet: sym5	73.12	84.29	71.91

When evaluating results, databases that include staged HIF signals exhibit lower accuracy than a trained and tested database that consists solely of modeled HIFs. There is no significant difference in accuracies based on the change of mother wavelet except for the ANN having a logistic function and 100 neurons. For this case, when a database that combines actual and modeled HIFs and utilizes the mother wavelet db4 was selected, accuracy decreased dramatically. While the logistic function showed better results for 1 Hidden layer and 100 Neurons, it became vice versa for tanh.

On the other hand, the results for the energy sum approach are shown in Table 4.3. ANNs trained with a database that includes actual and simulated HIFs have lower accuracy rates than the other two databases. There are apparent differences in accuracies based on the change of the mother wavelet. Furthermore, an ANN configured with two hidden layers and 50 neurons utilized sym5 as the mother wavelet, outperformed db4. In contrast, an ANN with one layer and 100 neurons trained with energies from db4 wavelet coefficients became superior to sym5. Additionally, the tanh function showed better results than the logistic.

Table 4.3 : ANN results when inputs are the sum of energy of coefficients for four levels.

Configuration	Actual HIF	Modeled HIF	Both
Activation: Logistic Layer: (100,1) Mother Wavelet: db4	82.45	79.86	59.75
Activation: Logistic Layer: (50,2) Mother Wavelet: db4	67.29	68.59	62.62
Activation: Tanh Layer: (100,1) Mother Wavelet: db4	93.25	92.56	88.33
Activation: Tanh Layer: (50,2) Mother Wavelet: db4	81.92	80.07	75.66
Activation: Logistic Layer: (100,1) Mother Wavelet: sym5	67.71	67.22	58.20
Activation: Logistic Layer: (50,2) Mother Wavelet: sym5	77.86	75.82	68.00

A similar investigation was conducted using SVM. Two different kernel functions were considered for each tested mother wavelet. Therefore, four outcomes for each trained database were obtained. The outcomes are displayed below in Table 4.4.

According to the findings, SVM with linear kernel function slightly outperformed SVM with RBF when trained with the standard deviations of wavelet coefficients. Nevertheless, all results achieved over 92% precision in distinguishing HIFs from other events when the standard deviation approach is considered.

On the other hand, implementing the energy sum approach resulted in reduced accuracy across all three databases examined. Moreover, when both actual and modeled HIFs were included in the database, the precision of HIF detection did not exceed 62%. Regarding kernel functions, no substantial differences were noted, although the results when the mother wavelet was sym5 were superior to those for the db4.

Table 4.4 : SVM results when inputs are standard deviation and sum of energy of coefficients.

Input Method	Database	SVM-RBF		SVM-Linear	
		sym5	db4	sym5	db4
Standard Deviation Approach	Actual HIF	91.94	94.27	94.31	93.78
	Modeled HIF	96.24	96.67	99.47	97.16
	Both	94.94	92.27	95.52	92.43
Sum of Energy Approach	Actual HIF	80.84	67.61	79.39	70.00
	Modeled HIF	80.29	66.43	79.88	71.45
	Both	61.28	61.30	61.24	61.43

4.4.2 Scenario 2: Results of noise-added signals

Under this section, the impact of noise was assessed using the same machine learning methods and input approaches. To achieve this, noise with SNR of 20, 40, and 60 dB was added to the existing current waveforms in the databases. This process generated new current signals with added noise. Afterward, these noisy signals were tested using the pre-trained algorithms. The prediction results of these methodologies were analyzed at a sampling rate of 100 kHz, employing the sym5 wavelet as the selected

mother wavelet for all experiments. Findings were presented in Table 4.5 and Table 4.6 for two distinct input approaches. The results indicate that the performance of algorithms declined with the increase in decibels of noise, regardless of the machine learning method or input approach used. The standard deviation of wavelet coefficients utilized as input to machine learning methods became superior to the energy sum approach for all cases. ANN achieved the highest accuracy rates when noise-added signals are considered.

Table 4.5 : Performance of ANN and SVM with noise when inputs are the sum of energy of coefficients.

Noise Level	ANN		SVM	
dB	(1,100)	(2,50)	RBF	Linear
0	58.20	68.00	61.28	61.24
20	45.87	41.71	34.05	38.77
40	43.40	37.82	29.45	32.73
60	32.01	35.20	25.14	29.29

Table 4.6 : Performance of ANN and SVM with noise when inputs are standard deviation of coefficients.

Noise Level	ANN		SVM	
dB	1,100	2,50	RBF	Linear
0	95.15	71.91	94.94	95.52
20	67.90	59.37	49.65	56.21
40	59.17	55.88	45.28	51.45
60	56.50	49.20	37.34	45.89

4.5 Discussion and Conclusion of Studied HIF Detection Methods

Machine learning algorithms were assessed for HIF detection, considering almost all non-HIF events and the influence of noise. The results indicated that the ANN

algorithm provided the best performance for noisy and noise-free signals. ANN method demonstrated the highest accuracy across various noise levels. The study revealed that using the standard deviation approach to feed machine learning methods outperformed the energy sum approach. Notably, when only modeled HIF cases were considered, both machine learning methods achieved greater than 99% classification accuracy with the standard deviation approach. However, it was observed that increased noise levels tended to decrease algorithm accuracy, regardless of the employed algorithm or input methods used.

Additionally, the selection of mother wavelets, specifically Sym5 and db4, did not significantly affect accuracy. Nevertheless, the Sym5 wavelet produced slightly better results for the database that included both modeled and actual HIF signals, suggesting that while the choice of wavelet may not be critical, it can still influence outcomes to some extent.

5. HIF LOCALIZATION STUDIES WITH ANN IN BRANCHED DISTRIBUTION NETWORK BY EVALUATING FEATURE SELECTION METHODS

Distribution utilities responsible for supplying electricity face challenges in maintaining a continuous power supply and serving end consumers. Locating faults in the distribution network is one major challenge since faults can cause long-duration disruptions in power supply. Therefore, effective fault localization techniques, particularly for high impedance faults, have become a key area of focus. Existing studies have primarily employed machine learning methods with features derived from signals. However, none have applied feature selection strategies or justified their preferred features by comparing them with different feature subsets. Moreover, it was found that test systems typically did not involve a significant number of buses and lateral branches. Additionally, the studies generally considered direct resistance with low ohm values as the representations of HIF to simplify the complexity of HIFs. Furthermore, methods relying on fundamental electrical parameters are less affected by noise, whereas techniques based on high-frequency components may be vulnerable [64]. However, numerous studies did not assess the impact of noise on the method's accuracy, even though they used features obtained with high-frequency information.

Therefore, the mentioned gaps are tried to be fulfilled by evaluating the performance of feature selection techniques in locating HIFs before offering a new localization approach in the thesis. Recorded current signals from the laboratory were used as fault data in the IEEE 34 bus distribution system in the simulation environment. The five most offered statistical features are utilized to train Artificial Neural Networks: standard deviation, variance, skewness, kurtosis, and the sum of energy content. These features are derived from extracted details and approximation coefficients in 7 levels from HIF current signals using Discrete Wavelet Transforms. Afterward, feature selection methods, which are wrapper, filter, and embedded methods, were employed to choose the best subset from the created 40 features. The selected subsets from each

selection method feed an ANN to locate HIFs. All statistical features utilized in the feature pool and the examined feature selection techniques are detailed in the following subsections, along with the simulation results.

5.1 Statistical Features

The extraction of meaningful features from data can be achieved through various methods. Statistical approaches are commonly employed for feature extraction to recognize patterns and attributes within the data. Based on the reviewed papers, it is evident that many studies have focused on utilizing only a limited number of statistical variables. However, the present study aims to encompass a comprehensive range of statistical variables. The statistical features that have been utilized are elaborated upon below.

Standard deviation: It (abbreviated as σ) is a measurement of how scattered the data (X_i) is in relation to the mean (μ). The majority of values are centered in one area. When the standard deviation is low, the data are centered around the mean, and when it is large, the data are wider. The formulation of the standard deviation is given below. N represents the number of the data points.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}} \tag{4.2}$$

Variance: It (σ^2) is the average of the squared differences of values in the data from the mean, which is the average or the most common value in the data.

Skewness (Sk): It is an indicator of a lack of symmetry. Data that is symmetric typically exhibits a skewness near zero, as the skewness for a normal distribution is zero. The deviation of each value from the mean is where the fundamental skewness begins. They are cubed and added together. Then this sum is divided by the cube of the root mean square standard deviation and number of data points. The formulation of skewness is shown below.

$$Sk = \frac{\sum_{i=1}^{N}(X_i - \mu)^3}{(N-1)\sigma^3} \tag{4.3}$$

Kurtosis (Kr): This metric measures whether the data has heavier or lighter tails compared to a normal distribution. The formulation of Kurtosis is given below.

$$Kr = \frac{\sum_{i=1}^{N}(X_i - \mu)^4}{(N-1)\sigma^4} \tag{4.4}$$

The sum of Energy Content (Ex): To distinguish the characteristic features of data with higher values, the square of the absolute value of data can be summed. The result of this process is called the sum of energy content. The formulation of it is presented below.

$$Ex = \sum_{i=1}^{N}(X_i)^2 \tag{4.5}$$

5.2 Feature Selection Methods and Methodology

5.2.1 Feature Selection Methods

Using an extensive set of features leads machine learning algorithms to overfit and increase computational cost. Moreover, all the features in the dataset may not be relevant. Generally, only a few variables in the dataset are practical, and the rest of the features are either redundant or irrelevant features. Utilizing these redundant features may decrease the performance of the algorithm. Feature selection methods come into the picture to avoid such negativity. The aims of feature selection are removing uncorrelated features and finding the best set of features from the data while reducing computational costs. In other words, this process is applied to reduce the computational load and enhance the result accuracy of the method by minimizing the input data.

The methods for feature selection can be divided into three categories by considering the presence of label information: supervised, unsupervised, and semi-supervised approaches. Since the study aims to estimate the precise location of HIFs, supervised feature selection methods were evaluated. These methods can be grouped into wrapper, filter, and embedded methods based on their selection strategies and integration with the learning algorithm [171].

Selecting a suitable feature selection technique relies on the distinctive characteristics and requirements of the dataset and the underlying learning task. Wrapper methods use a specific learning algorithm to assess iteratively the performance of features. On the other hand, Filter methods do not rely on any learning algorithm; instead, they determine feature significance based on data characteristics. Embedded methods

incorporate the advantages of both filter and wrapper techniques by merging feature selection with model learning and eliminating the need for iterative feature set evaluations.

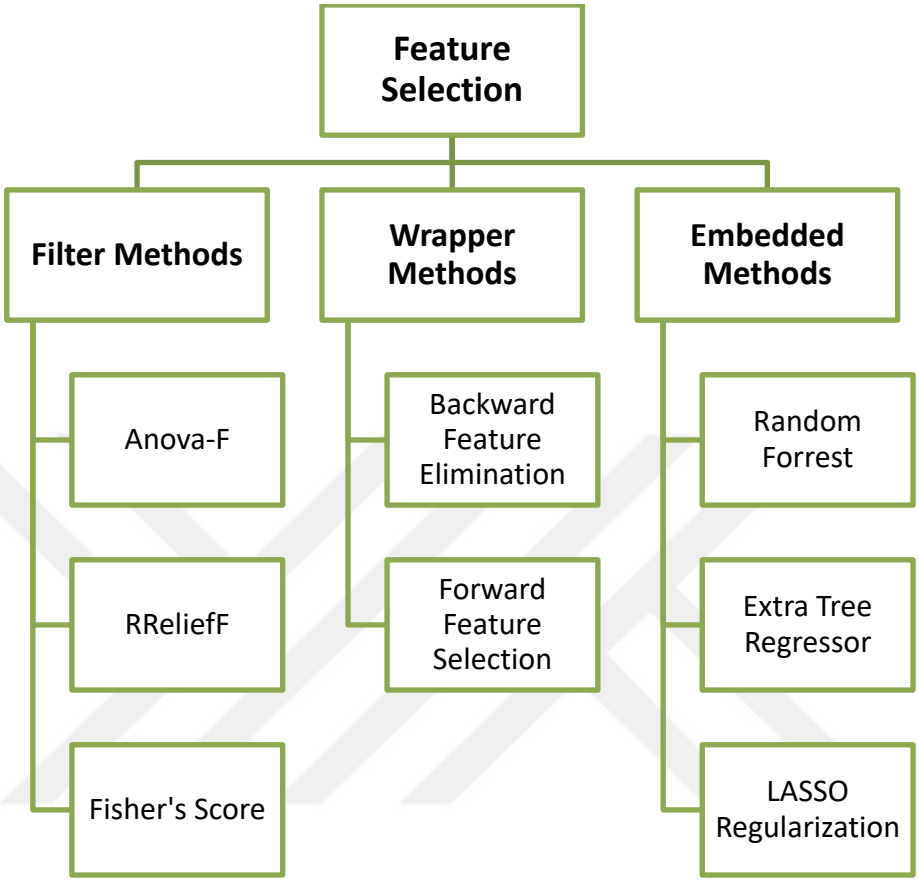


Figure 5.1 : Supervised feature selection methods covered in this thesis.

Supervised feature selection methods considered in the study are given in Figure 5.1. These feature selection methods were investigated based on the accuracy of their selections from the created feature pool.

5.2.1.1 Filter methods

Anova-F: The method examines each feature individually to assess its relationship with the response variable. It is used to measure the degree of linearity between the input feature and the output feature. A higher F-value suggests a strong linear relationship. Although it is a straightforward method for understanding how a feature is related to the response variable, it is important to note that this method is solely capable of catching linear relationships.

RReliefF: RReliefF is a widely used method in the Relief family that can be utilized for continuous regression problems. It conducts feature selection by establishing a

model that relies on the proximity of relevant features to other features within the same classes, as well as their distance from different classes [172].

Fisher's Score: Fisher's score is one of the most commonly utilized techniques for supervised feature selection. The theory behind it is based on the claim that high-quality features should provide comparable values for instances belonging to the same class while giving distinct values to instances from different classes.

5.2.1.2 Wrapper methods

Forward Feature Selection (FFS): This iterative method involves starting with the feature and testing it against the target features. Then, the different variable is selected based on how well it performs when paired with the initially selected variable. This procedure repeats until the specified criterion is met. It's important to note that the selected features may change in consecutive runs. Since there is no ranking in this wrapping method as the filter method, only features based on the selected number will be given at the end of the selection.

Backward Feature Elimination (BFE): This method performs in contrast to the FFS method. The technique removes one feature at each iteration. It starts from all features in the pool and continues eliminating features till the criterion is satisfied.

5.2.1.3 Embedded methods

Random Forest (RF): Random Forest is a popular tree-based feature selection method introduced by Tin Kam Ho and developed by Leo Breiman [173], [174]. It involves building numerous trees using random subsets of features by bootstrapping and then combining their results to find the best solution. Similar to filter methods, Random Forest determines the importance of features.

Extra Trees Regressor (ETR): It is derived from Random Forest and offered in 2006 [175]. It differs from RF in the way of selecting split nodes. The Extra Trees Regressor, also known as the Extremely Randomized Trees Regressor, chooses split points randomly while RF looks for the best split nodes.

Lasso Regularization: The Least Absolute Shrinkage and Select Operator (LASSO) is a type of regularization that minimizes the values of coefficients multiplied by the penalty factor [176]. It uses L1 regularization, which compels the coefficients of less

important features to become zero. This eliminates irrelevant features from the feature pool.

5.2.2 Methodology of the Study

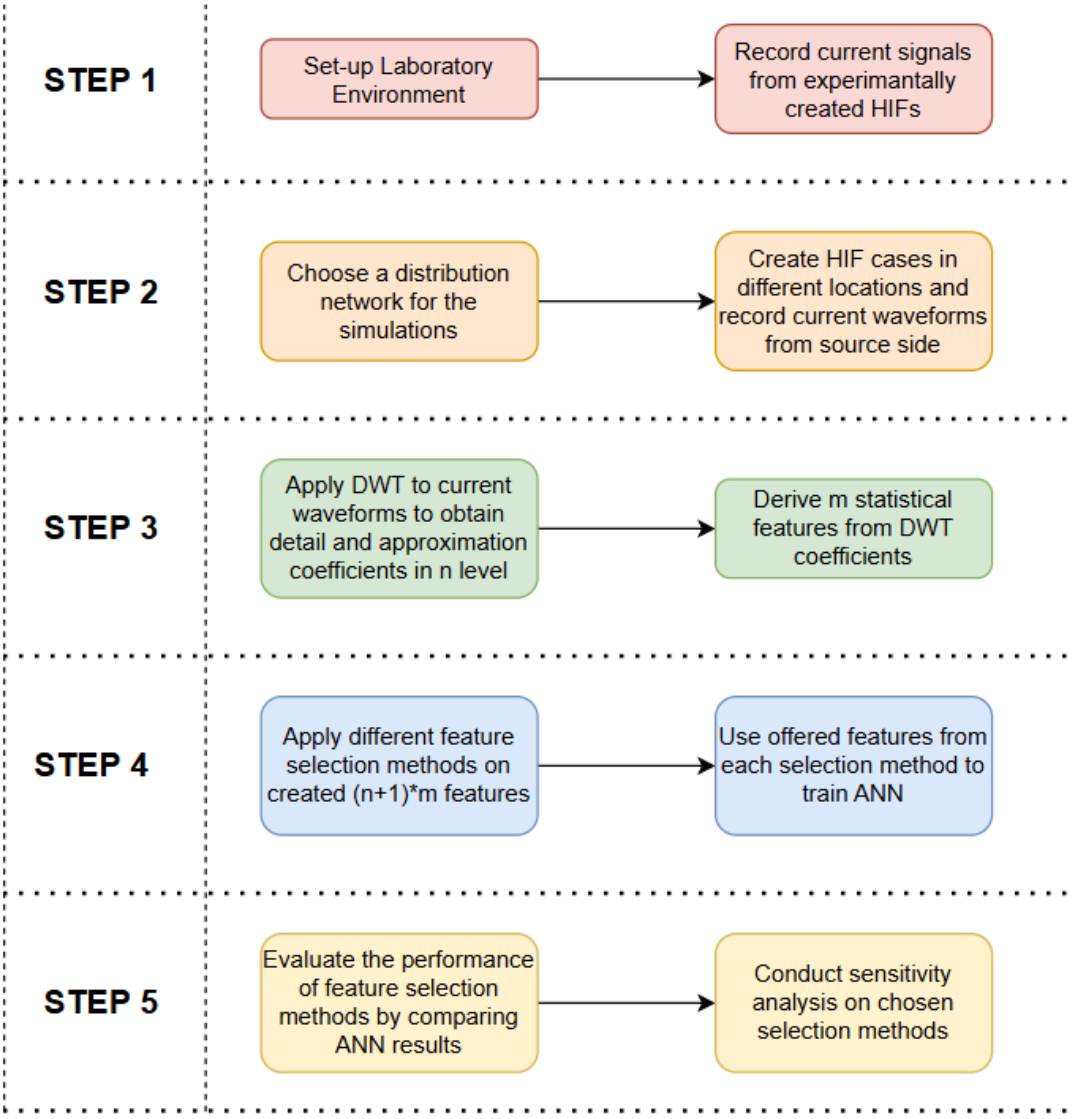


Figure 5.2 : Steps of Applied Methodology.

This subsection explains the methodology of the study. It outlines the flow of the studies conducted step by step. In the first step, necessary arrangements for the experimental tests were made in the High Voltage Laboratory at Istanbul Technical University. Next, HIFs were created and during the fault, current signals were captured for use in the simulations. Following this, a distribution network with lateral branches was selected and modeled to represent the actual system. The recorded signals were

then scaled, and HIFs were introduced as current sources in various locations to create cases.

After establishing the HIF cases, a DWT was applied to the current signals obtained from the simulations at an n-level. From each extracted coefficient, several (m) statistical features were derived. Subsequently, various feature selection techniques were implemented to determine the optimal subsets of features for use in the ANN. The effectiveness of these feature selection methods was assessed based on the results produced by the ANN. Moreover, additional assessments were conducted on the selected feature methods that provided the best outcomes. Figure 5.2 illustrates the structure of the methodology employed in the study.

5.3 Simulation Results and Discussion

Simulations were carried out in the IEEE 34 buses test benchmark radial distribution network. The network is modeled in a Matlab environment, and sampling rates of 100kHz were chosen on a 60 Hz base frequency. Since machine learning algorithms were chosen as the HIF localization method, numerous cases were created to train algorithms. These cases included different fault locations, fault resistances, and inception angles. HIFs were placed at all 34 buses and every 500 meters between sequential buses in 142 fault locations. The magnitudes of faults were randomly assigned between 5% and 15% of the nominal current of the network. The duration of cases was 150 ms, and the beginning time was changed in a cycle slot (0-90-180-270 degrees) for each simulated case. Moreover, 10 random fault magnitudes were assigned for each location. In other words, 40 cases were created for each fault location. Current waveforms for all three phases were recorded in the source bus during the simulations. Later, DWT is applied to current signals to extract detail and approximation coefficients in 7 levels. Five statistical parameters are calculated for each coefficient (D1, D2, D3, D4, D5, D6, D7, A7). At the end of the process, 40 features are obtained.

All the created simulations for training and testing data have been listed in Table 5.1. After that, some of the features among them are selected to train the machine learning algorithm. 70% of cases were used for training the algorithms, while 30% were used for testing. It should be noted that the method for pinpointing the HIF location assumes that the HIF has already been distinguished and identified.

Table 5.1 : Number of Simulated Cases in Terms of Related Event

Simulated Case Definitions	Number of Cases
HIF Placed on a Bus	1360
HIF Placed Between Sequential Buses	5680
TOTAL	7040

The mentioned feature selection methods were evaluated by testing and comparing them using the several variables they offered. An ANN having an activation function as a rectified linear unit (ReLU), lbfgs solver, 1 hidden layer, 100 neurons, and 500 epochs were considered. This structure is accepted as the base configuration in the study. The recommended features from each feature selection method were tested in the ANN with consecutive 250 times. Afterward, the average differences between the faulted location and the estimated location in terms of meters are calculated. In the following section results of selected features are presented to assess the effectiveness of various feature selection techniques. For better evaluation, it should be noted that ANN trained with 40 features yielded with 101.92 meter (m) estimation error.

5.3.1 Filter methods

The study involved calculating and ranking the importance of features for 40 features using filter methods. A varying number of suggested features were then tested to compare their effectiveness. In addition to evaluating the higher-ranked features with the ANN, the lower-ranked features were also examined to understand whether successful discrimination between relevant and irrelevant features was possible.

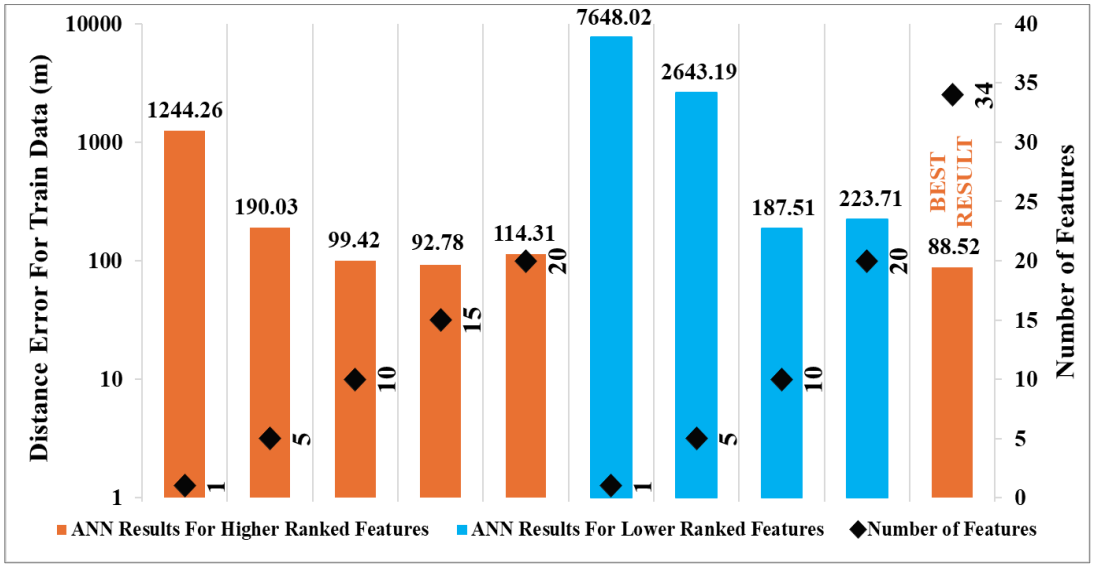


Figure 5.3 : ANN results based on the selected features from the Anova-F method.

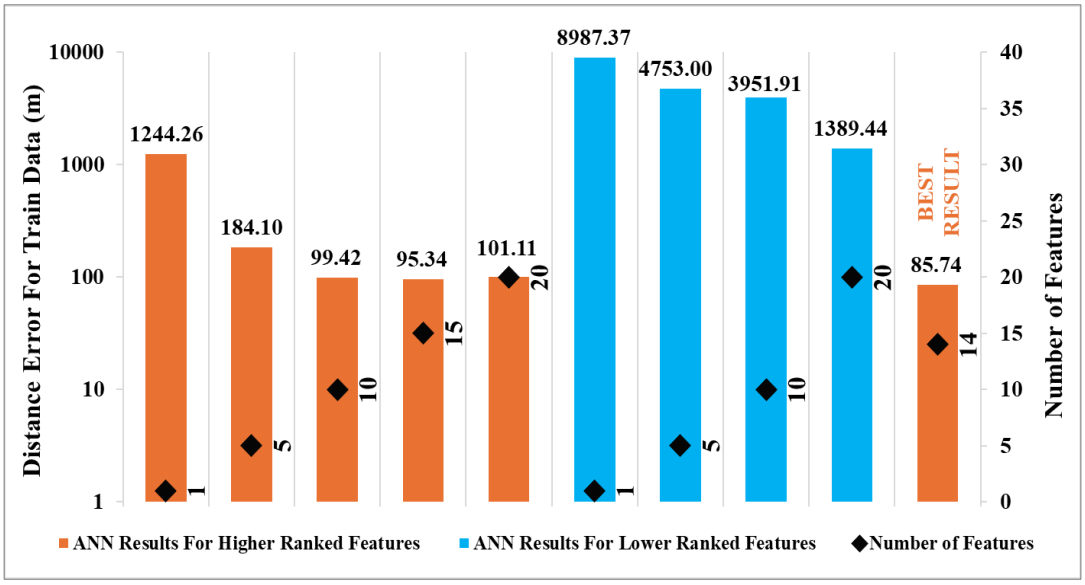


Figure 5.4 : ANN results based on the selected features from the RReliefF method.

The results of train data errors obtained after ANN trained with features from the Anova-F and RReliefF algorithms are presented in Figure 5.3 and Figure 5.4, respectively. Both approaches identified Kurtosis, obtained with approximation coefficients, as the top-ranked feature. While they differed after the second feature, similar results were achieved in the algorithms up to the first twenty features. However, the features chosen by the RReliefF algorithm result in lower errors when all tested higher ranked feature subsets are investigated. ANN trained with the last ten features gave a 3951.91 m distance error with the RReliefF, while a 187.51 m error was observed with the Anova-F features. Higher error values indicate that the RReliefF algorithm is more accurate than Anova-F in discriminating irrelevant features.

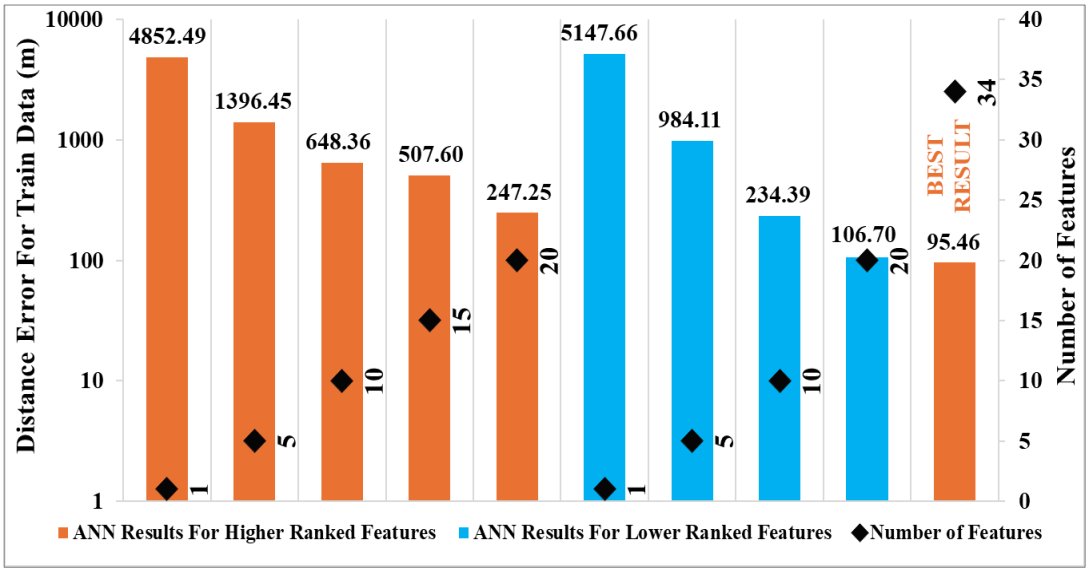


Figure 5.5 : ANN results based on the selected features from the Fisher method.

On the other hand, selected features using the Fisher method did not yield as good results as the previous two filter methods. The results for different feature subsets are illustrated in Figure 5.5. It can be observed from the figure that not only the most highly ranked features resulted in higher errors, but also the less prominent ones produced lower estimations.

5.3.2 Wrapped methods

Wrapper methods utilize learning algorithms in the feature selection process. Moreover, both FFS and BFE methods iteratively search through different subsets of features. Therefore, these methods are expected to produce better feature subsets. Each method has its pros and cons in terms of working principles. For example, FFS requires less computation time when evaluating fewer features, while BFE can eliminate most irrelevant features more precisely.

The results for FFS and BFE are illustrated in Figure 5.6. Not only are the results for the same number of features compared, but the best results are also compared. For up to 10 features, the results are relatively close to each other. However, while FFS identified 15 features as the best subset, BFE selected 13. Based on the results, FFS yielded 33% higher precision than BFE. The outcome is believed to be a result of the combination of features. Using more features to train ANN may decrease precision when one or more features are involved in the subset, but they may increase accuracy when fewer features are utilized.

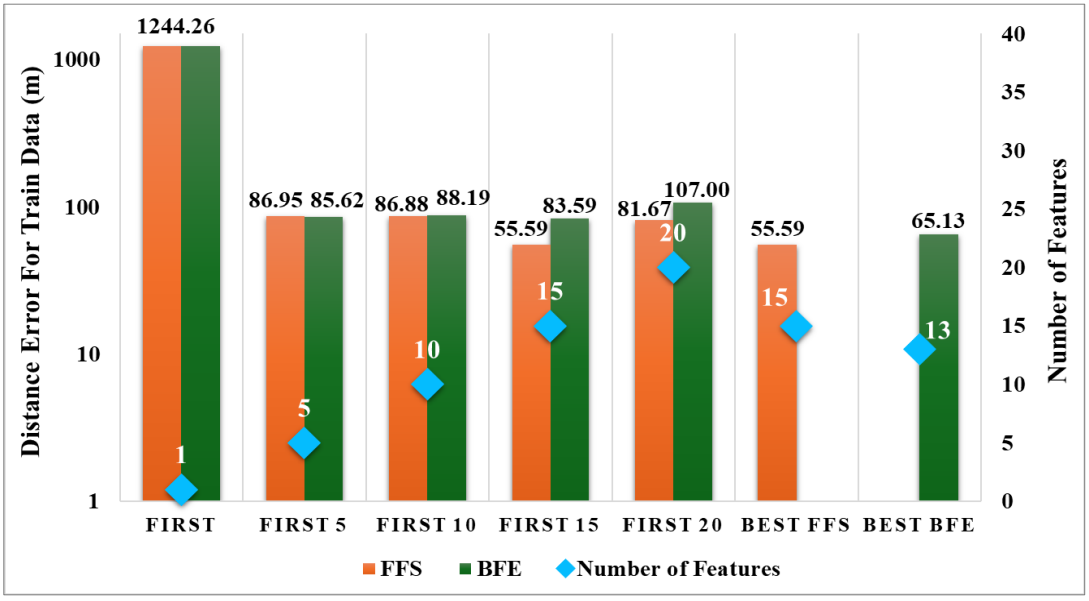


Figure 5.6 : ANN results based on the selected features from FFS and BFE methods.

5.3.3 Embedded methods

Embedded methods combine the strengths of wrapper and filter methods. They engage with the learning algorithm and provide feature sets without requiring iterative evaluations. This means less computation time compared to wrapper methods. Similar to filter methods, features can be scored to determine feature sets. Regularization and tree-based models, the most popular types of embedded methods, were tested in the paper.

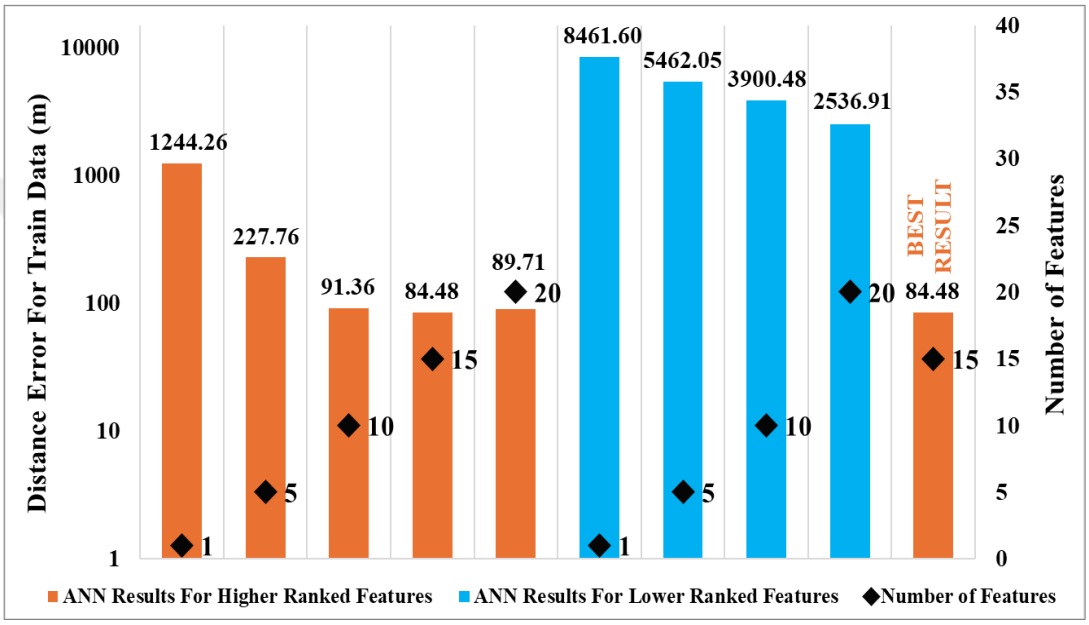


Figure 5.7 : ANN results based on the selected features from the Extra Tree method.

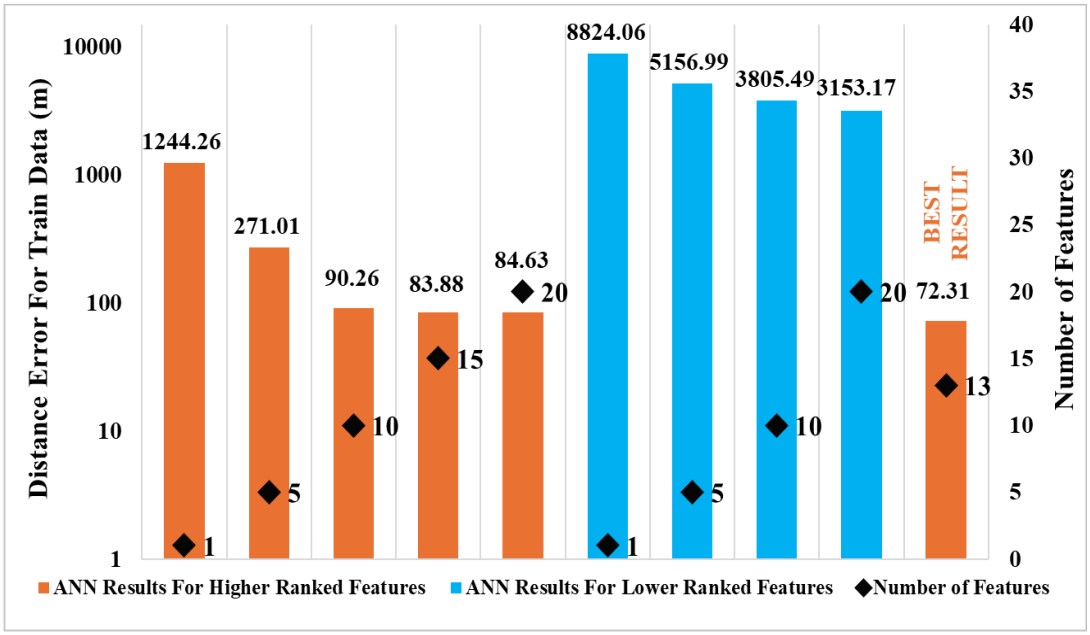


Figure 5.8 : ANN results based on the selected features from the Random Forest method.

The results for ETR and RF are presented in Figure 5.7 and Figure 5.8. As indicated in the figures, the average error associated with fault location estimation decreased with an increased number of features. Both algorithms demonstrated similar error distances, while they differed in the best results and number of features to reach these results. Random Forest outperformed the Extra Tree method in terms of accuracy of best result. Moreover, they displayed superior performance compared to filter methods. It is also worth noting that both algorithms effectively identified irrelevant features based on the less offered features. Furthermore, the algorithms have the potential to achieve higher accuracies by tuning their hyperparameters.

Different penalty terms are considered for Lasso regularization. However, it is observed that higher penalty terms often lead to the removal of a majority of features. Therefore, the maximum penalty term chosen is 0.1. Since the Lasso method set some feature coefficients to zero, not all feature subsets were thoroughly assessed for comparison. The results are demonstrated in Figure 5.9. Lasso regularization removed 31 features while offering 9 features with a penalty term of 0.1, and ANN trained with the remaining features exhibited lower errors than other penalty terms. Furthermore, when the same number of features, which is nine, are used to feed an ANN, the Lasso method outperforms all other methods.

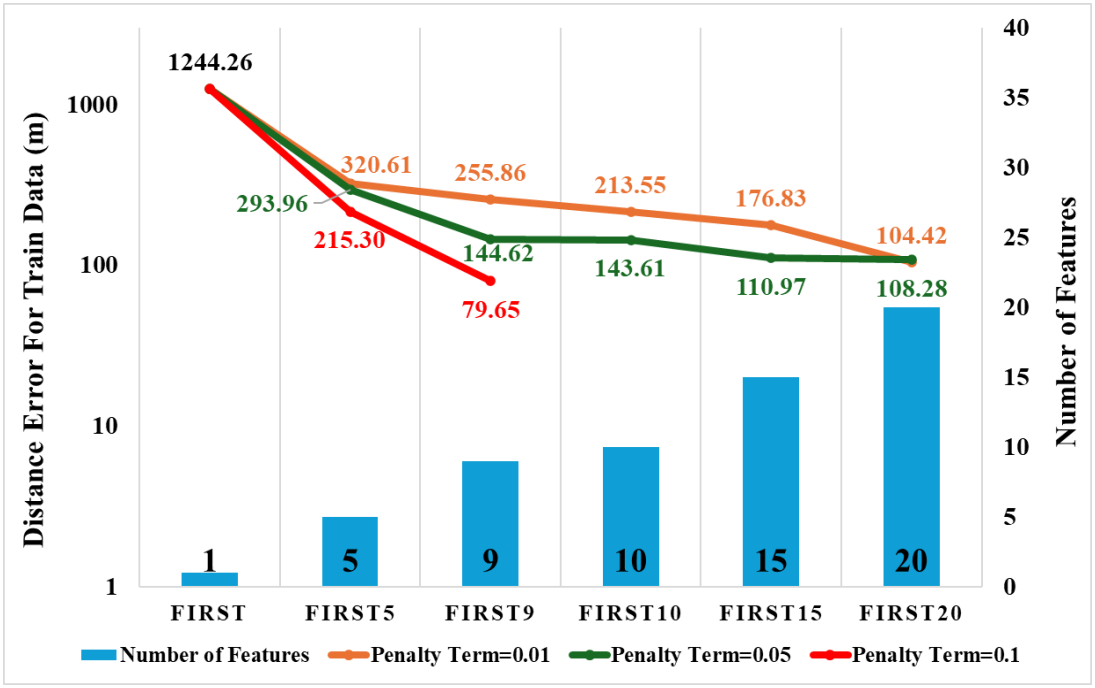


Figure 5.9 : ANN results based on the selected features from the Lasso Regularization method.

5.3.4 Sensitivity analysis on feature selection methods

Average accuracy errors of feature selection methods with their computation time and offered number of features to reach their best results are presented in Table 5.2. Wrapper methods yielded superior outcomes, while embedded methods produced the second-best results with lower execution time, as outlined in Table 5.2.

Table 5.2 : Evaluation of feature selection methods for their best feature subsets.

Feature Selection Method	Number of Features for Best Result	Train Data Error Distance (m)	Test Data Error Distance (m)	Execution Time of feature selection method (second)
Fisher	34	95.46	103.86	0.1199
Anova-F	14	88.52	94.33	0.0635
RReliefF	14	85.74	93.11	1.2552
Forward Feature Selection	15	55.59	60.30	137.6752
Backward Feature Elimination	13	65.13	69.10	149.1620
Random Forest	13	71.73	77.82	0.6777
Extra Trees Regressor	15	84.48	90.58	0.2102
LASSO Regularization	9	79.65	85.86	0.0648
No Feature Selection	40	101.92	111.09	-

Since the initial findings were obtained for an ANN configured with base parameters, further investigations were conducted using three distinct feature selection methods. Based on the initial results, wrapper methods demonstrated superior outcomes, and embedded methods yielded the second-best results with reduced computation time and fewer features. Therefore, FFS from wrapper methods, which produced the best result, was chosen as the first method. RF, the third-best result, and Lasso Regularization, ranking fourth with the lowest number of features, were selected as other methods from the embedded methods. The initial evaluation included comparing different numbers of epochs. Figure 5.10 shows the estimation errors of the training data for

FFS, Lasso, and Random Forest feature selection methods while epochs are 10, 50, 100, 1000, and 2000. It is evident from the figure that FFS generated the lowest errors in fewer iterations.

Additionally, the RF method exhibited a more pronounced reduction in the difference between calculated and actual fault location than others. Nonetheless, all methods yielded better results than using all features without applying any selection to them. Moreover, despite the Lasso method offering the lowest number of features, it achieved similar errors at higher epochs.

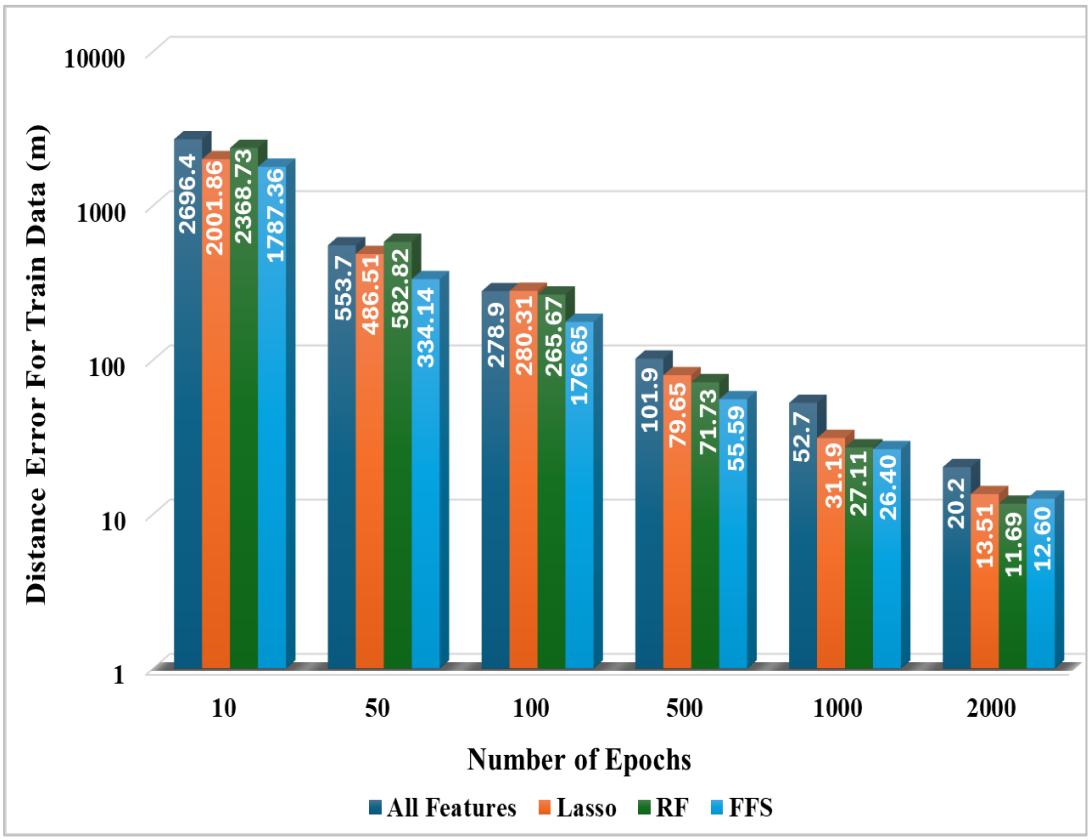


Figure 5.10 : Feature selection results with different epochs.

After the epochs assessments, various layer configurations and activation functions of ANN are investigated to understand the efficacy of the provided features. Findings for the different ANNs trained with the best feature subsets from the three feature selection methods mentioned earlier are presented in Table 5.3 and Table 5.4 in terms of mean and minimum estimation differences.

Upon evaluating Table 5.3, it is apparent that when the base configuration of an ANN is utilized, FFS features demonstrate superior accuracies compared to other configurations. Interestingly, a significant increase in errors occurred when two hidden

layers were utilized, along with 50 neurons. Additionally, the distance error doubled when features from the FFS method fed an ANN with two layers and 100 neurons instead of one layer. This dramatic increase was not observed with embedded methods, demonstrating their robustness to changes in the number of neurons and layers.

Table 5.3 : The results of ANN performance with chosen feature selection methods for different hidden layer configurations.

	Error Definition	Data	(Neuron=50, Layers=2)	(Neuron=100, Layers=2)	(Neuron=100, Layers=1) (Base)
FFS	Mean Distance Error (m)	Train	14543	123.39	55.59
		Test	14595	131.39	60.3
	Min. Distance Error (m)	Train	14009	68.25	38.8
		Test	14086	86.31	45.05
LASSO	Mean Distance Error (m)	Train	148.33	97.56	79.65
		Test	154.87	103.89	85.86
	Min. Distance Error (m)	Train	76.61	67.00	47.83
		Test	85.65	74.93	49.52
RF	Mean Distance Error (m)	Train	102.18	74.65	71.73
		Test	109.15	79.18	77.82
	Min. Distance Error (m)	Train	78.77	57.81	56.00
		Test	92.39	58.79	62.10

On the other hand, with the change of activation function, better estimations were achieved for embedded methods as indicated in Table 5.4. The average distance error when the Random Forest features feed the ANN configured with the Tanh activation function became similar to the results of FFS features. Moreover, features selected by both Lasso and RF methods led to decreased estimation errors and better alignment with the test data when considering the best ANN results.

Table 5.4 : The results of ANN performance with chosen feature selection methods for different activation functions in the hidden layer.

	Error Definition	Data	ReLU (Base)	Sigmoid	Tanh
FFS	Mean Distance Error (m)	Train	55.59	72.03	59.55
		Test	60.3	77.79	65.83
	Min. Distance Error (m)	Train	38.8	38.98	41.8
		Test	45.05	46.82	43.46
LASSO	Mean Distance Error (m)	Train	79.65	80.55	66.63
		Test	85.86	87.51	72.77
	Min. Distance Error (m)	Train	47.83	54.92	38.54
		Test	49.52	56.62	43.5
RF	Mean Distance Error (m)	Train	71.73	92.45	57.75
		Test	77.82	101.3	63.76
	Min. Distance Error (m)	Train	56.00	58.61	36.52
		Test	62.10	56.96	38.37

The study also examined how introducing noise affected the accuracy of the analysis. For this purpose, white noise was added to the recorded current signal to create new datasets. The estimation results of ANNs were then calculated for three different sets of features derived from feature selection methods. The impact of noise on the performance of the ANN is presented in Table 5.5.

The results showed that the average and minimum errors increased in the presence of noise. Moreover, the accuracy difference between training and testing data significantly increased when all features were considered. This result indicates an overfitting. However, this difference was insignificant when only offered features were utilized in the ANN.

Table 5.5 : ANN performances in base configuration with noise.

Feature Selection Method	Mean Difference Error (m)		Minimum Difference Error (m)	
	Train Data	Test Data	Train Data	Test Data
No Feature Selection	110.44	602.43	93.78	617.63
Lasso	242.93	345.25	221.86	330.87
RF	205.52	270.30	190.79	266.43
FFS	185.36	324.77	171.46	319.68

5.4 Discussion and Conclusion of Feature Selection Study in HIF Localization

Based on the literature, ANN, a popular machine-learning method, is widely used for addressing HIF localization problems. Thus, this study employed ANN to localize HIFs by examining various subsets of features derived from different feature selection techniques. The performance of these feature selection methods was evaluated based on the feature subsets they provide, which included varying numbers of the most relevant features. Additionally, subsets containing lower-ranked features, considered irrelevant, were also analyzed. Feature selection methods were categorized and examined within their classes before being compared to others.

Filter methods had the lowest computation time for ranking the importance of features since they evaluate each feature independently of any regressor. These methods generally provide initial insights into the dataset with favorable results. However, their independent evaluation of features, rather than feature subsets, can lead to less effective results when combining several highly ranked features. Additionally, some filter methods may not align well with the chosen machine learning algorithms, as demonstrated by the Fisher method outlined in this paper. A combination of lower-ranked features yielded better results in the Fisher method.

On the other hand, other evaluated filter methods, such as the Anova-F and RReliefF algorithms, extracted more reliable information on discriminating between relevant and irrelevant features. The RReliefF method demonstrated superior capability in filtering out irrelevant features compared to the Anova-F method. Using only the highest-ranked feature resulted in better accuracy comparing utilizing the 20 lowest-

ranked features in the RReliefF approach. Both algorithms reached their highest accuracy utilizing 14 features, exhibiting lower distance errors than those generated by the ANN when all 40 features were applied without any selection.

Wrapper methods are regarded as superior feature selection methods in offering the best subsets of features in terms of accuracy when the learning method and its configuration remain unchanged. However, they have longer computation times compared to other methods. Tested FFS and BFE techniques in this category yielded first and second-best results with the highest execution times.

Embedded methods are placed between these two approaches in time and error distance. The average results from the RF method closely matched those from the backward feature elimination method, even without optimizing their hyperparameters. Additionally, the Lasso method yielded the best results using the fewest features, whereas the others produced higher errors with the same top-ranked features. Although the feature subset obtained from ETR produced lower results compared to using all features, it cannot be stated that its best feature subset selection is superior to the Anova-F and RReliefF methods when considering distance error and execution time. When combinations of low-ranked feature subsets are examined, it is observed that the rankings from RF and ETR result in superior differentiation of irrelevant features compared to filter methods.

The effectiveness of three feature selection methods was analyzed using various configurations of ANNs, in addition to evaluating these techniques with a single ANN structure. The first evaluation involved changing the number of epochs. As the number of iterations increased, the distance errors decreased for all feature subsets, as expected. Up to 100 epochs, no significant differences in errors were observed between using features from embedded methods and using all features. Furthermore, the estimations of the ANN trained with RF features improved substantially with an increase in the number of iterations. In contrast, FFS produced lower errors than others with a smaller number of epochs. All tested feature selection methods performed better than using all features, but the RF method achieved the best results at the highest number of iterations.

After the epochs, the impact of changes in activation function and layer configuration was investigated. FFS features cause a significant increase in errors when there are

changes in the layer configuration within the ANN structure. It is believed that this is due to the dependency of wrapper methods on the learning algorithm and its configuration when choosing feature subsets. For the same reasons, the highest accuracy is accomplished when the base structure of ANN is utilized to test selected features. While the accuracies of the Lasso and RF methods varied with layer configuration, substantial changes were not observed, unlike in FFS.

On the other hand, both wrapper methods outperformed the FFS method when considering the best ANN results in different activation functions. The ANN structure with the tanh activation function led to the achievement of minimum errors. It's also important to note that optimizing the learning algorithm parameters could enhance the results of wrapper feature selection methods.

Along with the overall assessment of feature selection methods, their performance was investigated in noisy environments. The findings indicated that the accuracy of the algorithms decreased in the presence of noise. Moreover, overfitting occurred when the feature selection method was not considered. These outcomes highlight the importance of applying denoising techniques and feature selection methods to achieve improved results with noisy data.

6. PROPOSED HIF LOCALIZATION METHOD IN BRANCHED DISTRIBUTION NETWORK[2]

Fewer studies have focused on the localization of HIFs compared to their detection, as mentioned earlier. Although considerable research efforts have been dedicated to this issue, a universal solution that can be applied effortlessly and effectively in all situations for locating such faults is still lacking. Based on the studies conducted in Chapter 5, using a feature selection method may improve the accuracy of fault localization. However, these feature selection methods need to be applied to each specific network since the characteristics of different networks can vary, affecting the importance of each feature.

Therefore, a methodology that can identify the branch of HIFs in the distribution network and determine its location precisely is proposed within the thesis. Due to the lower accuracy of impedance-based methods compared to other techniques, which depend heavily on network parameters, a combined approach utilizing the traveling wave method and machine learning has been proposed. The goal of this combination was to address the drawbacks of both methods mentioned earlier. A discrete wavelet transform was applied to obtain traveling waves, while an Artificial Neural Network was selected as the base machine learning method due to its popularity in published HIF localization studies.

The proposed method has undergone numerous tests considering various inception angles and load variations to prove its robustness and effectiveness. The presence of shorter lines and multiple branches results in increased reflections and refractions, which negatively impact the accuracy of the studies. Thus, the proposed method was evaluated on IEEE 13 and IEEE 34 bus distribution networks using Matlab/Simulink

[2] This chapter is based on the following publication: **Baharozu, E.,** Ilhan, S., and Soykan, G. (2024). A Data-Driven-Based High Impedance Fault Location Method Considering Traveling Waves in Branched Distribution Networks, IEEE Access, 12, 186535-186546.

environments. Additionally, a comparative analysis was conducted on another test benchmark by comparing the results of the proposed method with other methods.

In the following, the basic concept of traveling waves and a detailed explanation of the proposed approach were presented. The details of the distribution system and several test cases are explained. Afterward, the findings from the simulation results are discussed in detail.

6.1 Theory of Traveling Wave Methods for Fault Localization

The concept of traveling wave techniques relies on high-frequency transient signals of voltage and current generated by faults. These transient signals propagate from the fault location in both directions toward the terminal bus. Specifically, by measuring the time difference between the first peak and later wave peaks at the terminal bus, along with information on the propagation speeds of both aerial and ground modes, one can calculate the fault distance. In the following sections, terminology related to traveling wave methods and the procedures employed in conventional traveling wave techniques were described.

6.1.1 Propagation speed

The velocity of a traveling wave propagates near the speed of light. Therefore, a high sampling frequency is required for the precise evaluation of waves. The propagation speed of a traveling wave along a power line is determined fundamentally by the network characteristics of inductance (L) and capacitance (C). The formulation of propagation speed is expressed quantitatively in (6.1).

$$v = \frac{1}{\sqrt{LC}} \tag{6.1}$$

6.1.2 Reflection and refraction of waves

When waves travel along a line and encounter a discontinuity, such as an open circuit, a short circuit, or a change in the line's impedance, the energy of the incident wave is distributed. A portion of the wave is transmitted (refracted) through the discontinuity, while the remainder is reflected toward the source. An illustration of these waves is presented in Figure 6.1. This behavior can be quantitatively described by the reflection

and transmission (refraction) coefficients, which are essential for analyzing wave propagation.

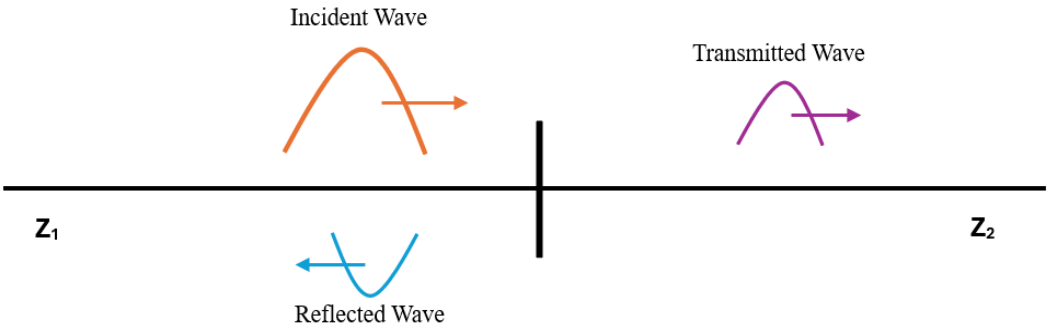


Figure 6.1 : Illustration of transmission and reflection waves at discontinuity.

The reflection coefficient (Γ) measures the strength of the reflected wave compared to the incoming wave, while the transmission coefficient (T) indicates the strength of the transmitted wave. The relationship between these coefficients is determined by the conservation of energy, which states that the summation of the reflected and transmitted waves must be equal to the incoming wave. The coefficients can be calculated as:

$$\text{Reflection Coefficient } (\Gamma) = \frac{Z_1 - Z_2}{Z_1 + Z_2} \tag{6.2}$$

$$\text{Transmission Coefficient } (T) = \frac{2Z_1}{Z_1 + Z_2} \tag{6.3}$$

Here, Z_1 and Z_2 represent the line's characteristic impedances on either side of the discontinuity.

6.1.3 Fault location with traveling wave method

As mentioned above, traveling wave methods rely on the propagation of the signal. They can be divided into two categories: active, and passive methods. The classification is determined based on how the wave pulse is generated. While the traveling wave pulse occurred by fault is used in the passive method, the pulse is injected in the active method to obtain waves and find the fault location [177]. In both cases, high-frequency transient signals called waves propagate along the line and reflect at terminals and the fault location. By analyzing the arrival times of these waves and calculating their propagation speed, it becomes possible to determine the distance to the fault. Based on this approach two distinct applications of the traveling wave method exist depending on the number

of measurement devices employed [178]. The single-ended method utilizes only one measurement device located at the source of the network. Conversely, the double-ended method requires multiple measurement devices placed on different buses. Therefore, the double-ended method uses several synchronized devices to locate faults whereas the single-ended methodology eliminates the requirement for communication between measurement devices, resulting in reduced costs. This technique involves detecting the arrival of the first traveling wave induced by a fault and recognizing the wave's reflection from the site of the fault. Afterward, the time interval between the original wave and the reflected wave is used to pinpoint the exact location of the fault.

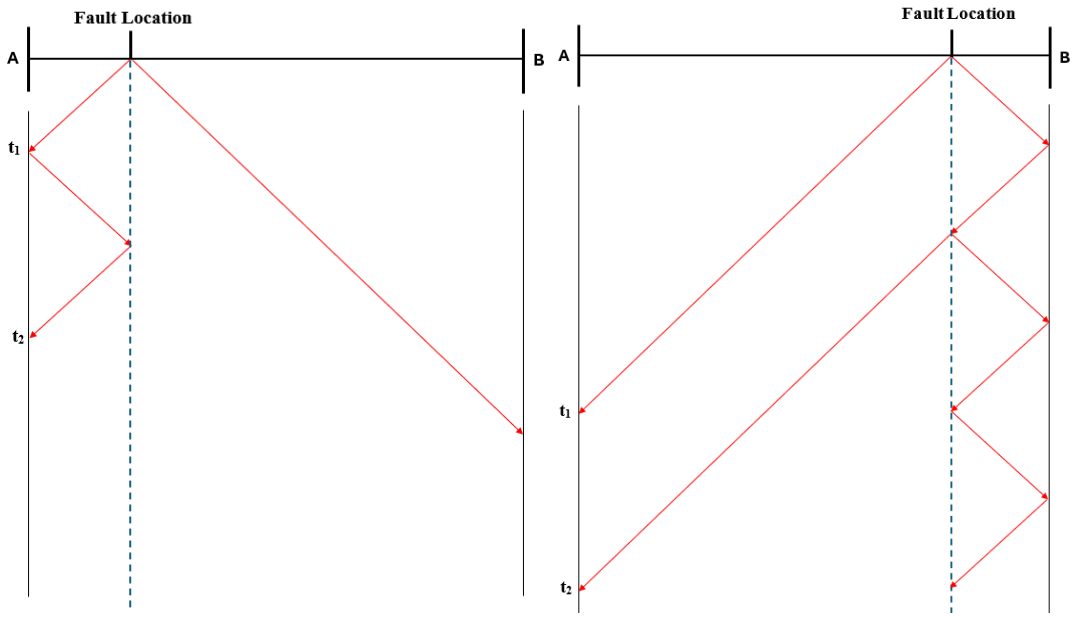


Figure 6.2 : The Bewley Lattice diagram for fault located in (a) the First and (b) the Second half of the line.

The Bewley Lattice Diagram for a fault located in the first and second half of a line is illustrated in Fig 5. Based on the fault location, calculation equations for fault distance change. When faults occurred at a distance shorter than halfway along the line, the equation used to determine the fault location is presented below.

$$X = \frac{1}{2} v\tau(t_2 - t_1) \tag{6.4}$$

In this formulation, v represents the speed of the wave, τ denotes the inverse of the sampling frequency, t_1 indicates the time of the first wave, and t_2 refers to the time of the reflection wave to reach the measurement point from the fault point.

When faults occur at a distance longer than halfway along the line, the equation used to determine the fault location changes to below.

$$X = Len - \frac{1}{2} v\tau(t_2 - t_1) \tag{6.5}$$

Here, Len represents the total length of the line.

6.1.4 Clarke transformation

Since electromagnetic coupling between conductors of three-phase lines results in different propagation speeds, the arrival times of waves used to calculate fault location cannot be determined precisely from phase signals [179]. Therefore, a modal transformation is employed in traveling wave methods to address this issue by decoupling phase signals into aerial and ground modes. Clarke, Karrenbauer, and Wedepohl transformations are commonly applied in traveling wave methods to obtain aerial and ground modal values [180].

This thesis utilizes the single-ended passive traveling wave method with the Clarke Transformation technique. The equation for Clarke Transformation [181] is provided below in (6.6).

$$\begin{bmatrix} I_\alpha \\ I_\beta \\ I_0 \end{bmatrix} = \frac{2}{3} \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix} \tag{6.6}$$

Where Iα and Iβ are alpha and beta aerial mode components and I0 is the ground mode component.

6.1.5 Signal processing in the time-frequency domain

Traveling wave methods utilize high-frequency signal elements to identify the location of faults. Therefore, techniques revealing information regarding the variations in frequency over time are required for traveling wave approaches. In this regard, The Short Time Fourier Transform (STFT) and the Wavelet Transform could be used since both define signals in the time and frequency domains. Discrete Wavelet Transform for signal analysis is used in this thesis. Since it is explained in previous sections, it is not detailed in this subsection once more.

6.2 The Proposed Method

Using traveling wave and ANN methods, a localization method for HIFs on the distribution network is proposed in the thesis. This method consists of five main steps. Here is a description of each step:

Step 1: Creating an initial dataset

Firstly, the distribution network model with frequency-dependent line models should be created in the simulation environment. Then, on this model, HIFs should be created with a modeling technique at different locations in the distribution network, and current waveform data should be recorded for each fault condition from the substation point.

Step 2: Processing initial dataset

In Step 2, Clarke transformation is applied for the measured current signals to obtain alpha aerial mode components. Afterward, a one-level discrete wavelet transformation was applied to signals to extract the time-frequency transient components on aerial-mode traveling waves that occurred due to HIFs. While Figure 6.3 shows the current signal shapes related to the aerial mode of the traveling wave signal during HIF, the part of transient components of this signal obtained by DWT are given in Figure 6.4.

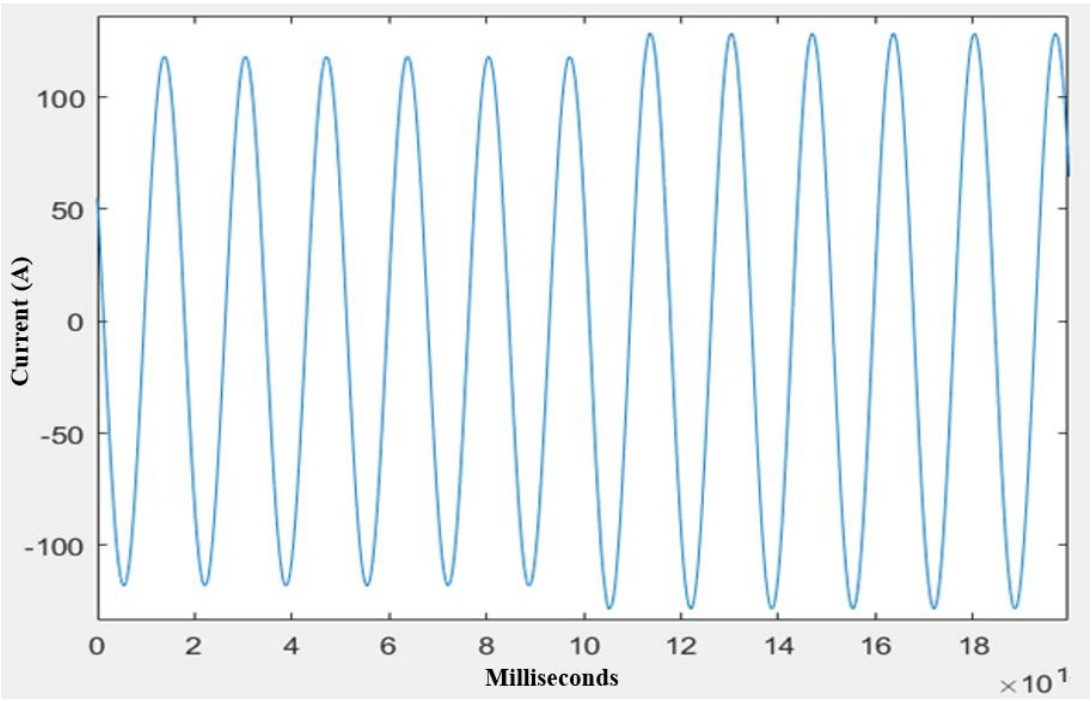


Figure 6.3 : Aerial mode of faulty phase.

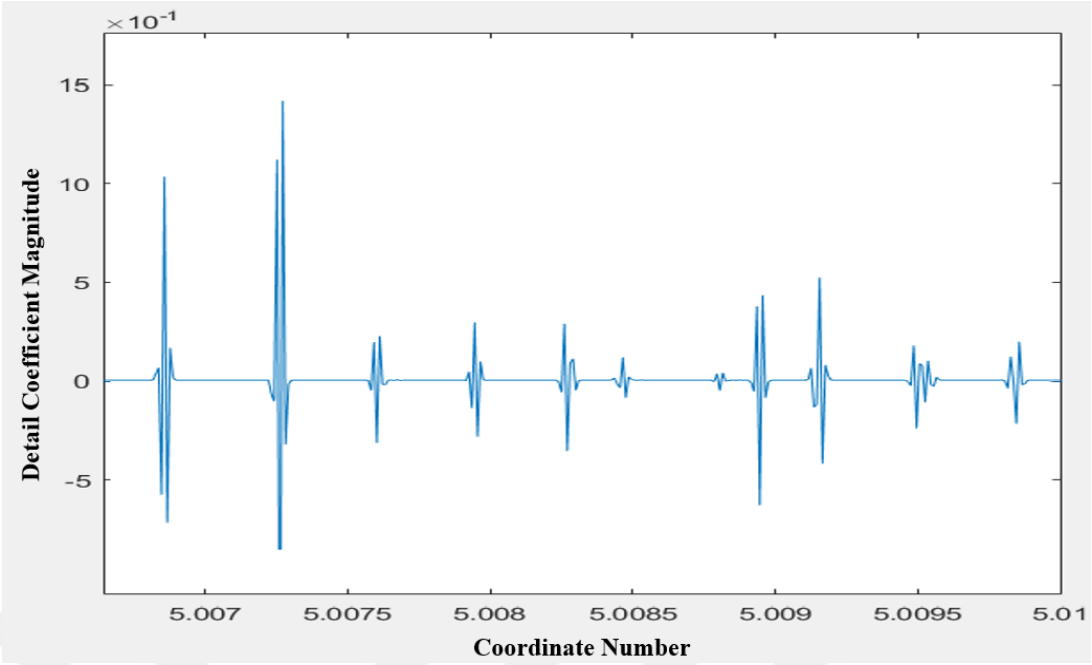


Figure 6.4 : Transient components of aerial mode signal after DWT is applied.

Step 3: Selecting and numbering traveling waves

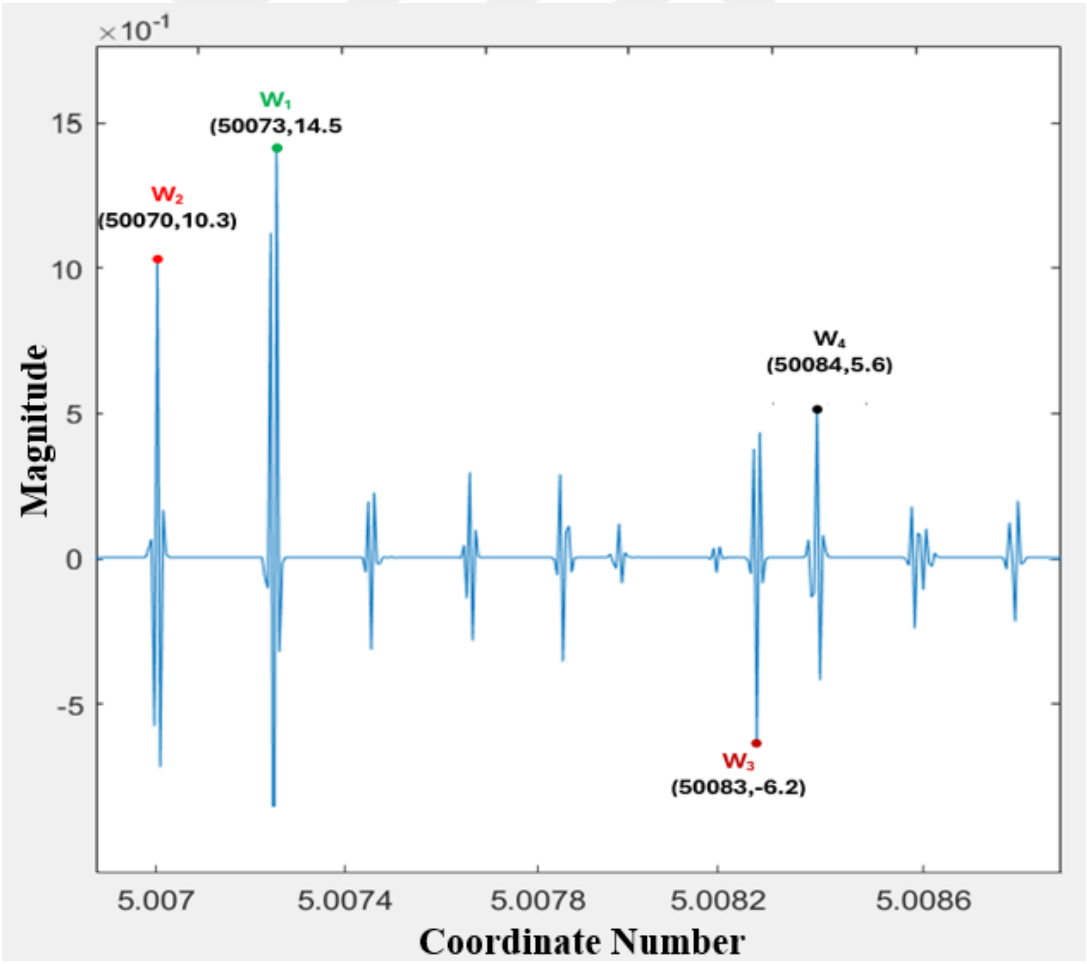


Figure 6.5 : Selected waves with their coordinate and magnitude information in DWT applied aerial mode of sample faulty signal.

In branched distribution networks, a high number of traveling waves is generated after a HIF due to reflections and refractions. Identifying each reflected and refracted traveling wave is quite challenging, so the proposed method offers to select a specific number of traveling waves to determine the fault location. Since the generated traveling waves depend on the total number of branches in the distribution network, the number of waves to be selected is taken as one less than the total number of branches. Once the number of waves to be selected is determined, they are chosen from the signals obtained by DWT based on the magnitudes of the wavelet coefficients at both polarities. The wave with the highest magnitude is selected as the first wave, followed by the second-highest, and so on, in descending order, until the desired number of waves is selected. In addition, the coordinate values of selected waves are kept in a CORD array. Figure 6.5 illustrates the numbering of selected waves in a faulty current signal in a network with five branches. Here, the coordinate numbers of four illustrated waves are recorded and assigned to the CORD array.

Step 4: Preparing input arrays for ANN

This step involves creating a binary representation of the traveling waves in the form of an array called POS. This array will be used as input in the training process of an ANN to determine the branch and location of the fault. The phases and array equations of this step are given below.

1. Before filling the POS array, construct an additional array DIF, using the elements of the CORD array.
2. The elements of the DIF array are computed by subtracting each selected wave's coordinate number from the initial wave's coordinate number.
3. Create an array of zeros, called POS, based on the highest computed element of the DIF array to account for all selected waves.
4. Use the elements of the DIF array to fill the elements of the POS array with one. If the value of an element in the DIF array is equal to the index of the POS array, assign one to the corresponding element in the POS array.

$$DIF(i) = CORD(i) - CORD(1), \qquad for\ 1 \leq i < n \tag{6.7}$$

$$POS(j) = \begin{cases} 1, & if\ j == DIF(i)\ for\ 1 < i < n \\ 0 & \end{cases}, \ for\ 0 \leq j \leq \max_{1<i<n}\left(DIF(i)\right) \tag{6.8}$$

Step 5: Applying ANN to find the fault location and its branch

In this step, the ANN undergoes the training phase using a supervised learning algorithm that involves backpropagation. For this purpose, the arrays generated in step 4 are submitted to train the neural network. After the training stage, ANN will be utilized. When the HIF is detected in the network, a POS array will be created and given to trained ANN to accurately determine the fault location and identify the branch of the network where the fault occurred.

Figure 6.6 illustrates the structure of the proposed method for the HIF localization scheme.

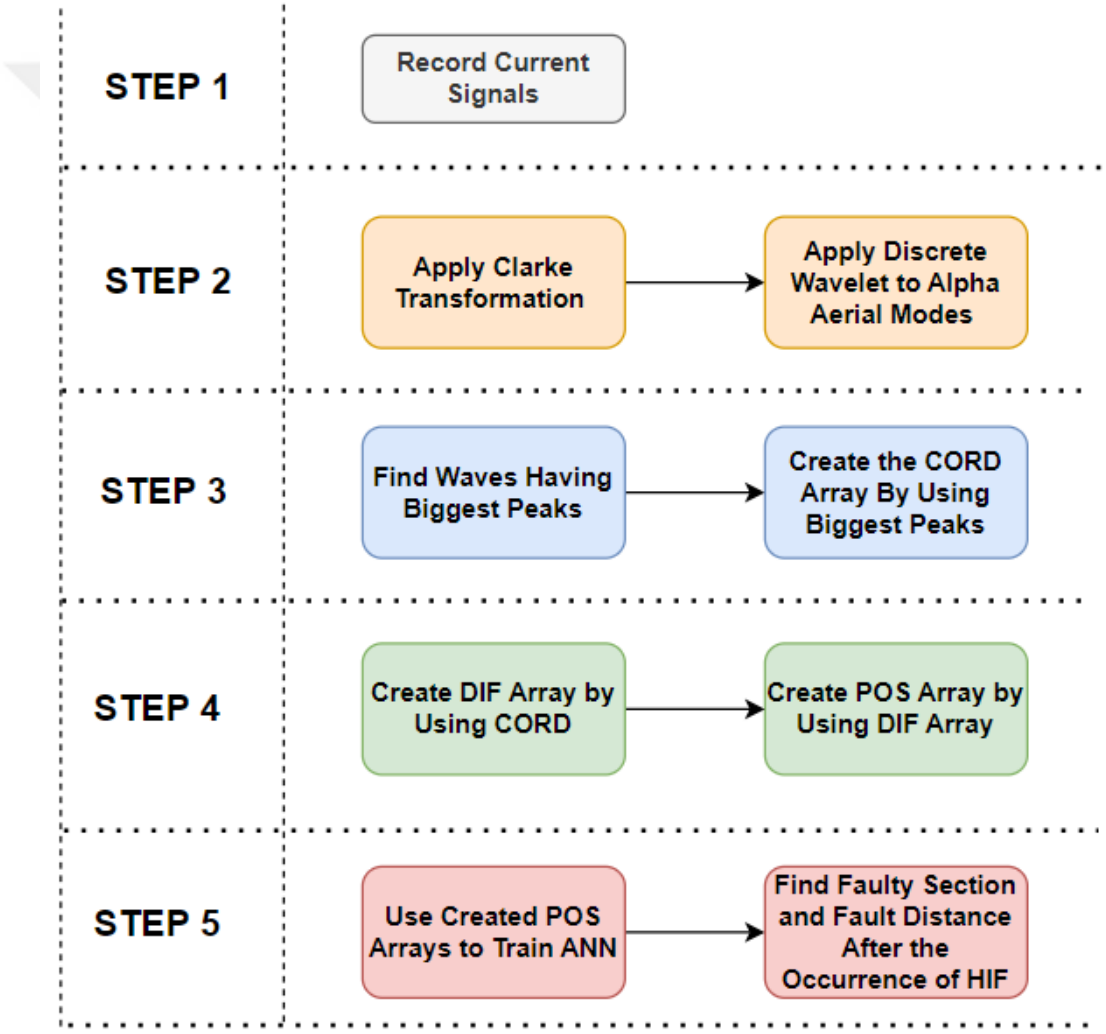


Figure 6.6 : Structure of proposed method.

6.3 Simulation Cases

The proposed method has been tested on three different distribution networks. The primary simulations were conducted on the IEEE 13 bus test benchmark radial

distribution network, as shown in Fig. 6. Besides, the IEEE 34 bus network was used to test the method's effectiveness on a larger network. In addition to these systems, 14 bus distribution network was considered for comparative analysis. These selected test systems are heavily unbalanced and consist of several branches. Moreover, the length of the branches and network are considerably small in the IEEE 13 test benchmark for applications of traveling wave methods. The test networks were simulated using Matlab/Simulink software with frequency-dependent line models. Since the IEEE 13 and IEEE 34 bus networks consist of sections shorter than 100 meters, the minimum sampling rate should be 3 megahertz (MHz). However, to avoid overlapping in reflections and refractions while choosing peak waves, a sampling rate of 10 MHz was chosen on a 60 Hz base frequency. Even though smaller sampling rates can be used for larger networks, advancements in technology have significantly enhanced the capability to capture signals at sampling rates of 10 MHz in field applications. High-performance microprocessors now facilitate the effective acquisition of data at these elevated frequencies.

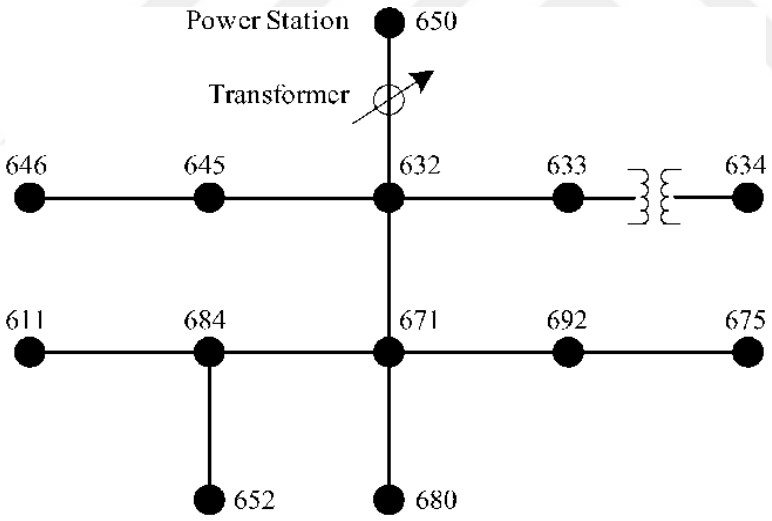


Figure 6.7 : IEEE 13 bus test benchmark.

In the realm of power system analysis, high impedance faults have been a topic of significant interest for researchers. Over the years, several models have been proposed to represent the distinctive characteristics of HIFs. The electric arc, which occurs in the HIFs, generates specific features in the HIF's current waveform. Furthermore, the electric arc leads to asymmetry between the positive and negative half-cycles of the signal and results in non-linear behavior between voltage and current at the fault point [92]. Therefore, the study used the HIF model presented in [29] to represent the mentioned characteristics.

Voltage sources and limits of resistor values were taken from [51] for the IEEE 13 bus test system, and from [182] for the IEEE 34 bus test system. The following items show the situations considered in the simulations:

- The method for pinpointing the HIF location assumes that the HIF has already been identified.
- Datasets were created by applying HIFs in the network with randomly assigned fault resistances and different inception angles.
- HIFs are placed on each bus and every 500m between the buses if the distance exceeds 500m.
- The magnitude of resistance was selected by considering uniform distribution while adhering to the limits specified in the references [51] and [182].
- The starting time of the fault was changed 16 times in a cycle slot (each 22.5 degrees) for each fault location to create various inception angles. The ANN in the proposed method was trained using 70% of all input data, while the remaining 30% was used for testing.

The proposed scheme was tested on different networks with several sensitivity analyses to demonstrate its robustness. The accuracy of faulty sections, the difference between actual and estimated fault distances, and their error percentages were evaluated in the study. Subsequently, the ANN was run 50 times for each analysis, and the averaged results were presented below in subtitles. For the analysis of the results, section accuracy was calculated by (6.9), and errors were calculated by (6.10) and (6.11):

$$Accuracy\ (\%) = \left[\frac{Number\ of\ Accurately\ Identified\ Sections}{Total\ Number\ of\ Cases}\right] x\ 100 \quad (6.9)$$

$$Error\ (\%) = \left[\frac{Calculated\ Distance - Actual\ Distance}{Actual\ Distance}\right] x\ 100 \quad (6.10)$$

$$Error\ (m) = [Calculated\ Distance - Actual\ Distance] \quad (6.11)$$

The steps of the method are illustrated with an example to enhance the comprehensibility of the proposed methodology before presenting the simulation results. After capturing the current waveforms on the source side in step 1, a Clarke transformation is applied to the signal. This results in the aerial modes of the fault

signal, as depicted in Figure 6.8. Then, the DWT is applied to the aerial mode current signal to obtain the traveling waves in step 2. Afterward, the first four waves with the highest magnitudes are selected from the DWT-processed aerial mode signals, as presented in Figure 6.9. The initial wave peaks at point 414699, while the second, third, and fourth highest waves peak at points 414707, 414744, and 414817, respectively. These values are used to create the CORD array in step 3, as demonstrated in Figure 6.10. By subtracting the first element value from each of the subsequent elements within the CORD array, the DIF array is created using the procedures outlined in step 4. Based on the CORD array, element values of the DIF array are assigned to 0, 8, 45, and 118 as shown in Figure 6.11. Then, the size of the POS array is set to 118, and a value of 1 is assigned to the array indexes of 0, 8, 45, and 118 to signify particular points of interest within the dataset. Figure 6.12 and Figure 6.13 provide visual representations of the constructed POS array. This array and additional POS arrays generated from remaining fault waveforms are then utilized to train an ANN with label information of faults' distance and section. Later on, trained ANN was utilized to identify faulty sections and estimate the distance to the fault from the source.

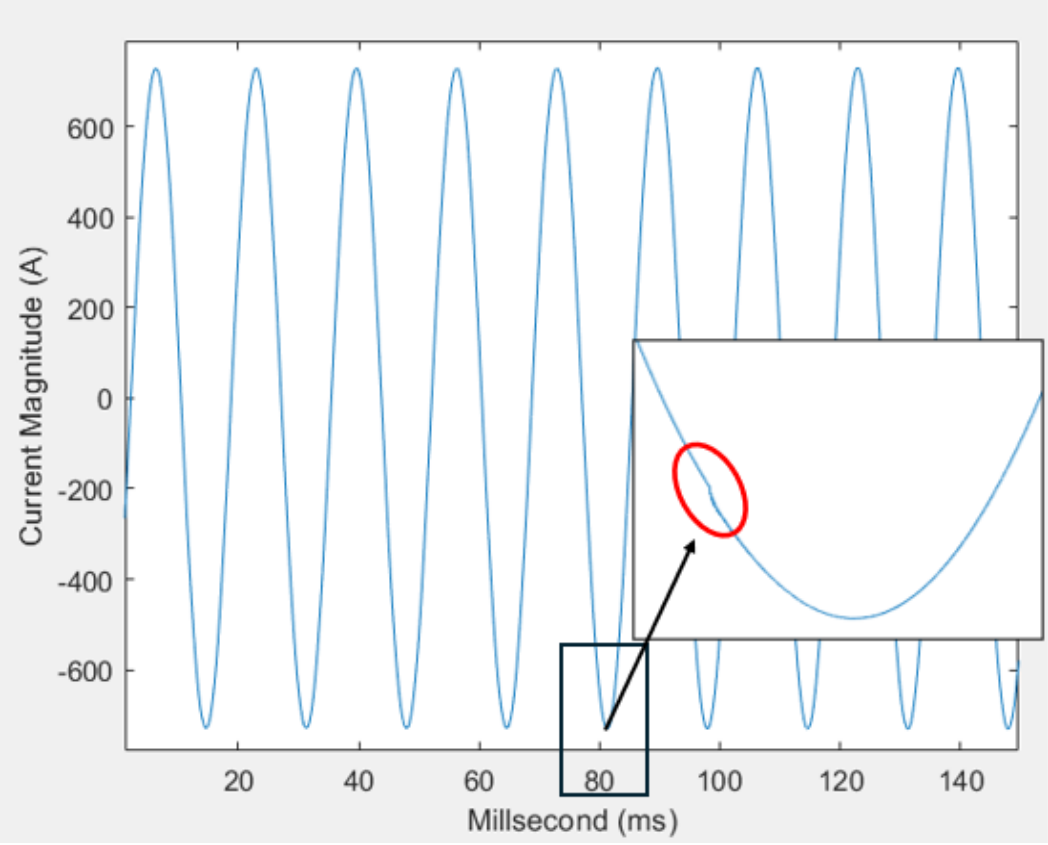


Figure 6.8 : Alpha aerial mode of faulted current signal.

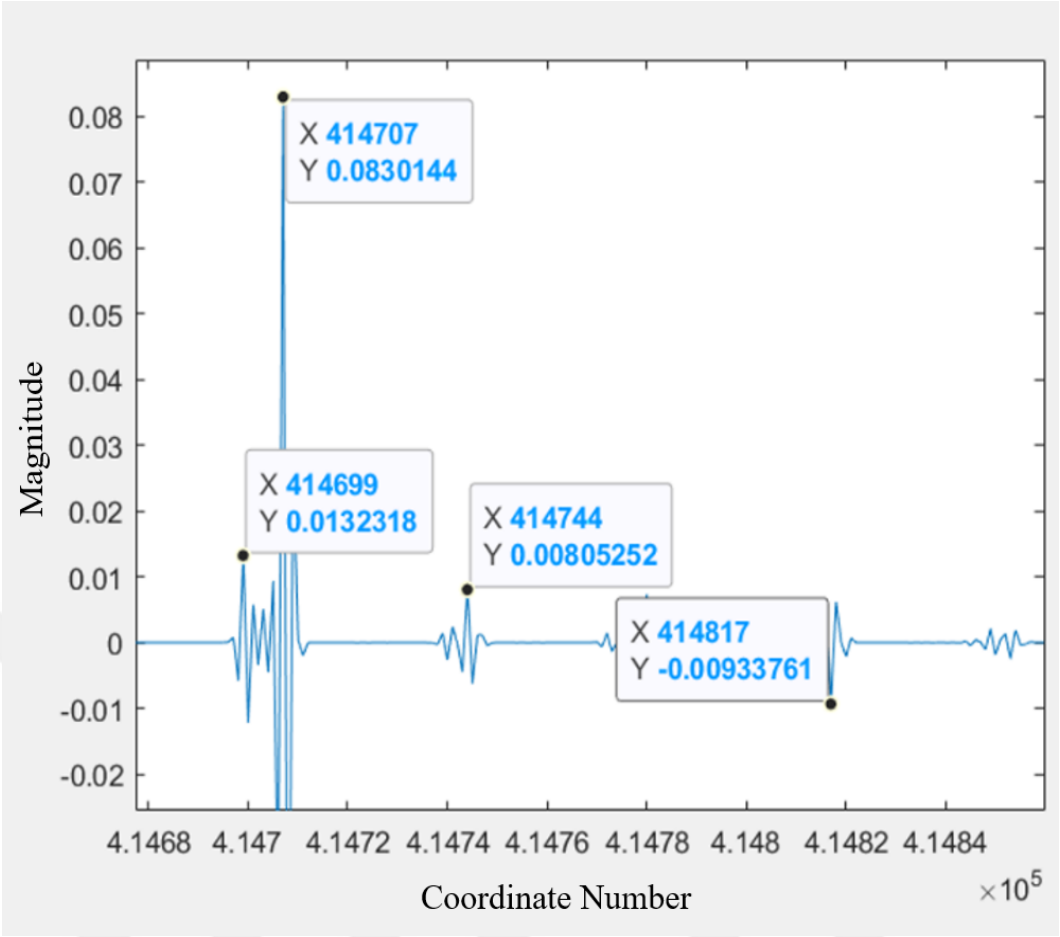


Figure 6.9 : DWT applied alpha aerial mode of faulted current signal and selected four waves having the highest magnitudes.

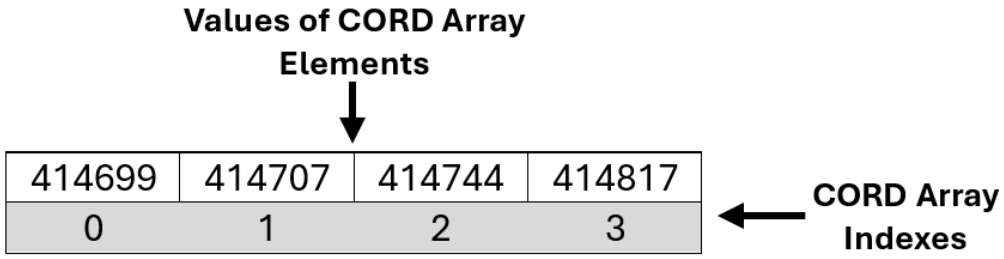


Figure 6.10 : Created CORD array based on selected waves.

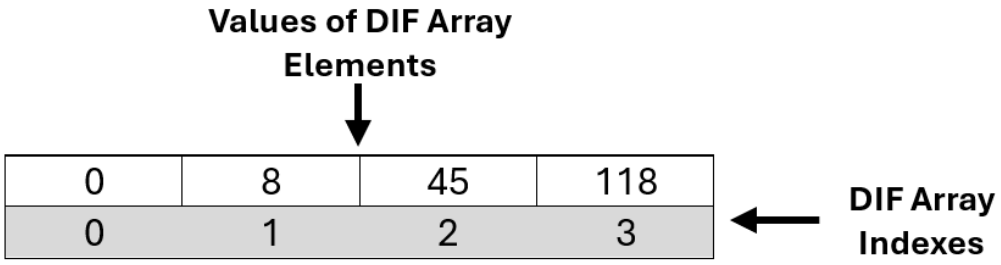


Figure 6.11 : Created DIF array based on CORD array.

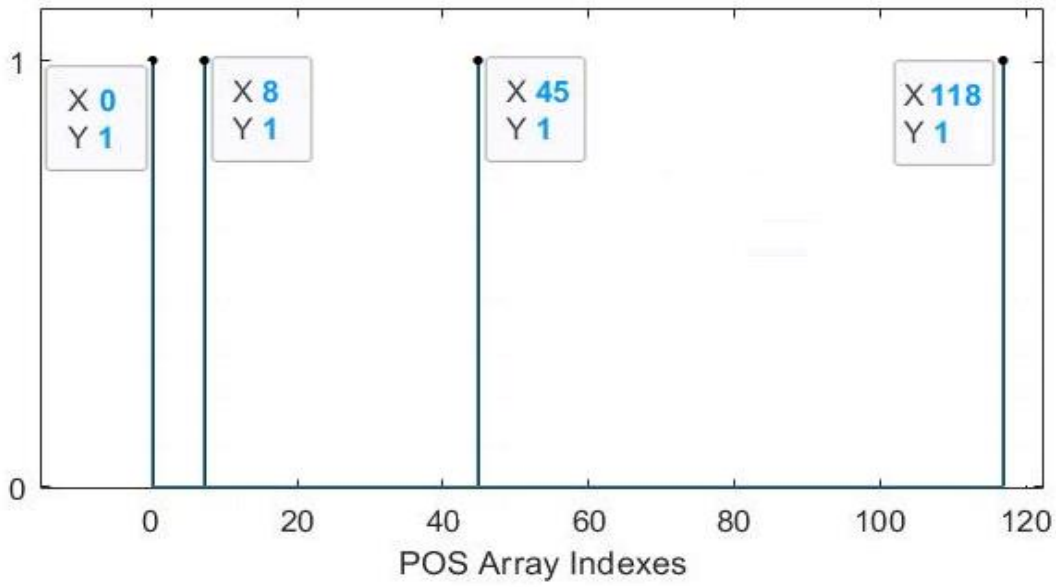


Figure 6.12 : Illustration of the created POS array from the DIF array.

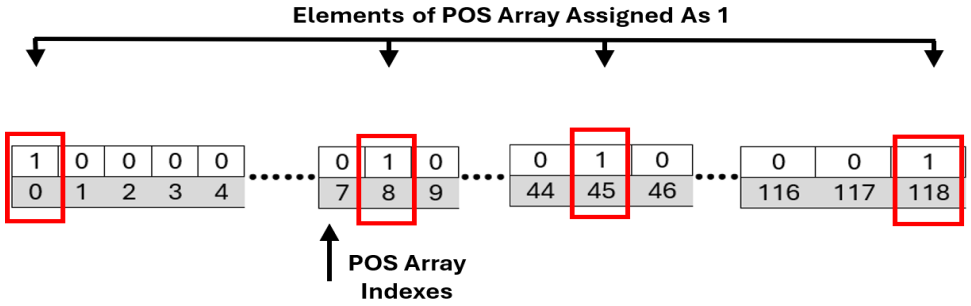


Figure 6.13 : One of the created POS arrays from traveling waves to fed ANN.

6.4 Simulation Results

6.4.1 Sensitivity analysis on IEEE 13 bus test system

The effects of different mother wavelets were evaluated in this part. In the beginning, ANN was considered one hidden layer with 50 neurons and 100 epochs while the loading of the network was 100% to make comparisons more clearly. Since the wave in overhead lines travels near the speed of light (3 x 10^8 m/s), if the only traveling wave method is employed instead of the proposed approach, the minimum accuracy would be around 30 meters with a 10 MHz sampling rate. By calculating the velocity of the traveling wave based on the line parameters in the tested networks, the minimum accuracy achieved could be a couple of meters less than 30 meters. Nevertheless, the proposed method can accomplish higher than the 30 m error, regardless of the mother wavelets used.

This study evaluates three distinct families of wavelets: Daubechies, Symlet, and Biorthogonal at various orders. The selection of these mother wavelets is made by their reported precisions, as documented in the research [183]. Furthermore, different faults will primarily impact different sub-ranges of frequency [184]. Therefore, an appropriate wavelet should be selected for power system transient studies depending on the waveform shape of the examined signal. Given that the objective is to identify peaks within transient components, it is anticipated that wavelets characterized by sharp terminations and minimal oscillation will yield superior results. After testing several orders of mother wavelets, the most promising ones from their classes are presented in Table 6.1. The average outcomes in Table 6.1 show that Bior 5.5 outperforms the other tested mother wavelets. Moreover, when considering only the best result of Bior 5.5 instead of the average results, the distance errors were obtained as 0.0104 m and 0.0083 m for train and test data, respectively.

Table 6.1 : Results of the proposed algorithm for different mother wavelets in the IEEE 13 bus test system.

Mother Wavelet	Accuracy of Faulty Section	Location Error of Train Data		Location Error of Test Data	
	Perc. (%)	Distance (m)	Perc. (%)	Distance (m)	Perc. (%)
Db4	70.31	11.0474	0.8560	20.7244	1.6562
Sym5	68.82	4.4168	0.3424	26.4214	2.1690
Bior 5.5	76.56	0.0166	0.0016	6.8526	0.6061
Bior 6.8	81.48	3.9342	0.3052	7.8902	0.6908

The effectiveness of the proposed method was assessed through the use of different ANN configurations, while Bior5.5 served as the mother wavelet, and loading was 100 %. The results presented in Table 6.2 indicate that the accuracy of the proposed scheme improves as the number of neurons in the hidden layer and maximum iterations (epochs) increase, particularly for identifying faulty sections. Although the configuration parameters of ANN could be further optimized through a systematic tuning process, it is essential to note that all tested configurations resulted in acceptable performance outcomes.

Table 6.2 : Results of the proposed algorithm for different structures of ANN in IEEE 13 bus test system.

ANN Structure	Accuracy of Faulty Section	Location Error of Train Data (m)	Location Error of Test Data (m)
Number of Neurons: 10, Maximum Iteration: 100	36.41	0.0100	16.7442
Number of Neurons: 50, Maximum Iteration: 100	76.56	0.0166	6.8526
Number of Neurons: 50, Maximum Iteration: 500	99.38	0.0007	6.9694

It is crucial to consider the varied loading conditions in distribution networks to see how they may impact the performance of fault localization methods. Specifically, methods that rely on extracted features from fault signals, such as harmonic information, RMS values, and phase differences, may exhibit instability under different loading conditions. Even though traveling wave methods are immune to such changes in the network, the proposed method was tested under 100%, 50%, and 30% loading conditions. ANN was considered one hidden layer with 500 neurons, while Bior5.5 is the mother wavelet. Results of the different loading are presented in Table 6.3.

Table 6.3 : Results of the proposed algorithm for different loading in the IEEE 13 bus test system.

Loading	Accuracy of Faulty Section	Location Error of Train Data		Location Error of Test Data	
	Perc. (%)	Distance (m)	Perc. (%)	Distance (m)	Perc. (%)
100%	99.38	0.0007	0.00006	6.9694	0.5909
50%	99.61	0.0004	0.00003	5.4238	0.5168
30%	99.45	0.0005	0.00004	6.0895	0.5419

The proposed method utilized machine learning methods independent of any specific algorithm. Since the results of the proposed method combined with the traditional ANN were satisfactory, more conventional learning algorithms were examined for comparison. The goal was to demonstrate the robustness and effectiveness of the proposed methodology by evaluating its performance with less advanced learning

techniques. Therefore, machine learning methods, including Support Vector Machine, Decision Tree, Gaussian Process Regression, and K-Nearest Neighbors (KNN), were evaluated. Based on the findings outlined in Table 6.4, ANN yielded superior results to others. Additionally, the proposed technique consistently produced acceptable error margins from distribution power utilities regardless of the machine learning method tested. These results support the suitability of the proposed approach for other machine learning and deep learning methods.

Table 6.4 : Results of proposed algorithm for different machine learning methods.

Machine Learning Methods	Accuracy of Faulty Section (%)	Location Error of Train Data (m)	Location Error of Test Data (m)
SVM	99.44	<10e-3	20.1273
DT	99.38	<10e-3	13.1379
GPR	98.65	<10e-3	89.9422
KNN	96.67	7.0965	34.7367
ANN	99.38	<10e-3	6.9694

6.4.2 Performance of the method in IEEE 34 bus test system

The method's robustness was tested with the larger network in the study. For this aim, the IEEE 34 bus test system was utilized. An analysis was conducted with ANN using varying numbers of neurons, with Bior5.5 as the mother wavelet, 100 epochs, and 100% network loading. The method's performance is detailed in Table 6.5.

Table 6.5 : Results for IEEE 34 bus test system for different numbers of neurons.

Number of Neurons	Accuracy of Faulty Section	Location Error of Train Data		Location Error of Test Data	
	Perc. (%)	Distance (m)	Perc. (%)	Distance (m)	Perc. (%)
500	95.44	9.9746	0.0174	29.689	0.0515
50	95.51	9.7155	0.0169	42.5568	0.0738

According to the results indicated in Table 6.5, the method presented in the study demonstrated high precision and accuracy when applied to the IEEE 34 bus test

system. On the other hand, fewer neurons slightly reduced the accuracy of the estimated fault locations.

6.4.3 Performance of the method in extreme conditions

The performance of the method might be less effective in networks having extreme conditions such as overloading. To show the proposed method's applicability and robustness, a loading of 150% was applied to tested IEEE networks to evaluate extreme conditions. Such loading causes a reduction in the ratio of fault current to nominal loads and subsequently results in fewer disturbances in the signals due to fault. Therefore, it is a good case for testing the method's effectiveness. Findings from the test results are shown in Table 6.6. The outcomes indicate that the proposed method establishes high accuracy in distinguishing faulty sections and estimating fault distance even under extreme loading conditions.

Table 6.6 : Results of the proposed algorithm for extreme loading conditions in IEEE 13 and IEEE 34 bus test systems.

	Accuracy of Faulty Section	Location Error of Train Data		Location Error of Test Data	
Network	Perc. (%)	Distance (m)	Perc. (%)	Distance (m)	Perc. (%)
IEEE 13 Bus	99.11	0.0017	0.00016	8.3657	0.6485
IEEE 34 Bus	95.43	9.9725	0.0183	34.571	0.0825

6.4.4 Comparative analysis in 14 bus test system

This section presents a comparative analysis of the proposed methodology with previously published methods to demonstrate its efficacy. ANN was considered one hidden layer with 50 neurons and 100 epochs while the loading of the network was 100%. The tests were conducted under identical conditions employing the network (Figure 6.14) disclosed in [185]. Subsequently, the outcomes were compared with those of [185], [133] and [150]. In Figure 6.15, 1M refers to a single measurement, and 3M stands for three measurements. In other words, results in the method [150], while using three different measurements to locate the fault, were represented as 3M. According to the results presented in Figure 6.15, the proposed method displayed

significantly lower errors than the compared methods. When considering only one measurement, the superiority of the proposed method became even more evident.

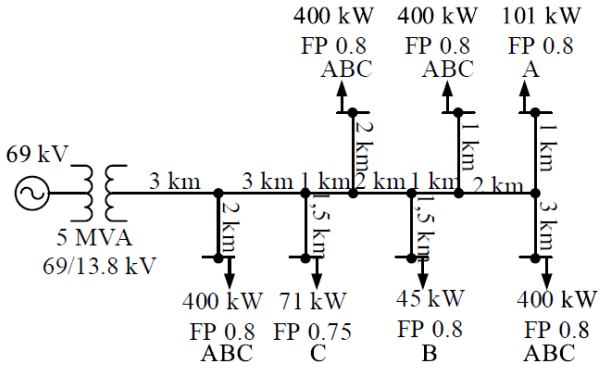


Figure 6.14 : Small test system used for validation of the proposed method [185].

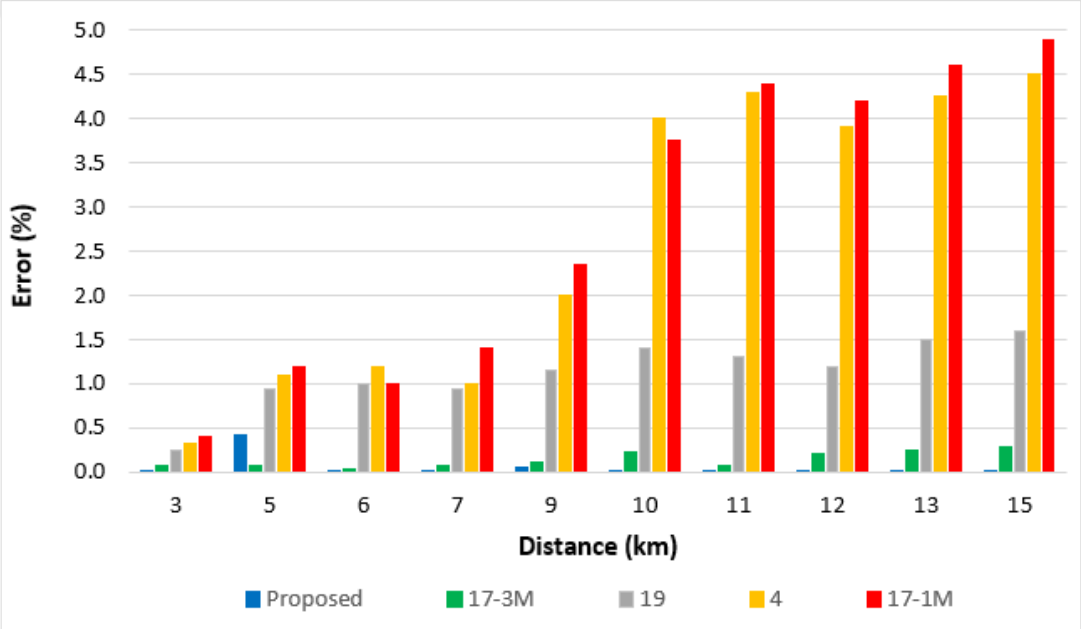


Figure 6.15 : Comparative analysis results.

6.5 Simulation Outcomes

The thesis proposes a new methodology to identify the location of high impedance faults in distribution networks accurately. The proposed method comprises two main steps by using both traveling waves and ANN. The first step involves identifying several traveling waves with higher magnitudes from current signals recorded at the substation and creating an input array based on these waves. In the second step, the method uses ANN to compute the distance of the fault from the source and determine the related section of the fault.

Thanks to the proposed methodology, the fault distance and identification of the faulty branch were determined using a single machine learning method. The performance and robustness of the method were tested on IEEE 13, IEEE 34, and 14 bus test systems, considering several factors such as load imbalance, changes in resistance, and variation in inception angle. The outcomes of all tested scenarios indicate satisfactory performance, with an error rate below 1%. Furthermore, the proposed method achieved higher precision compared to the direct use of the traveling wave method at the same sampling frequency.

The dependability and efficacy of the proposed methodology were investigated by utilizing other less advanced learning techniques instead of ANN. The results indicated that the proposed method yielded acceptable error levels, regardless of the machine learning technique. These findings validate the applicability and efficiency of the proposed method across different machine-learning approaches. Additionally, utilizing more advanced and recent learning algorithms within the proposed method can lead to more precise and accurate prediction results.

Furthermore, a comparative analysis was performed on the 14 bus test system, demonstrating the exceptional performance of the proposed methodology. All these findings can be interpretable as a significant advancement in the field, providing a more efficient approach to locating HIFs in power systems.

7. CONCLUSIONS AND RECOMMENDATIONS

Traditional protection equipment can detect various faults; however, high impedance faults often go undetected due to their low fault currents. These faults occur when a broken live conductor comes into contact with the ground or when surrounding vegetation leans against a live conductor. In both cases, a high-impedance path is created, resulting in barely noticeable changes in the power signals. As a result, HIFs can last for long durations without detected. Additionally, the arcing during such faults can cause severe risks, including the potential for fire and injury to a person who touches a broken conductor while it is still energized. For these reasons, researchers have concentrated on HIFs to develop definitive methods for identifying and locating all instances of these faults. Despite these efforts, a universal solution is still lacking. As a result, the thesis searched for an economical and practical solution, particularly for the localization of HIFs, while also improving the comprehension of HIFs.

The main contributions of the thesis are summarized as follows,

- The thesis emphasizes the importance of comprehensively understanding the characteristics of high impedance faults to enable accurate detection and location. Therefore, high-voltage experiments were performed at the distribution level, focusing on analyzing faulty signals derived from various types of HIFs.
- An extensive review of existing literature concerning HIFs was conducted to classify the various methodologies utilized in this field. This review aimed to highlight the strengths and weaknesses of different approaches while identifying significant gaps in the current understanding and research related to HIFs.
- Different distribution networks were utilized to evaluate the performance of detection and localization methods, ensuring they reflect real-world conditions with branches and unbalanced loads. Additionally, simulations of distribution

events that replicate the characteristics of HIFs were carried out to validate the efficacy of the detection algorithms.

- The thesis involved an evaluation of two different published HIF detection methodologies and their features within machine learning techniques, comparing the results to verify the effectiveness of methods by considering with and without experimental HIF signals.
- The gap in the existing HIF localization studies concerning the lack of feature selection strategies and comparative analyses between various feature subsets is found. Thus, the thesis investigates the effects of different feature selection methods on HIF localization and evaluates these methodologies based on their error distance.
- A novel approach integrating the traveling wave method with an artificial neural network was proposed to determine the fault distance and identify the faulty branch using a single ANN. The effectiveness and robustness of the technique were tested on IEEE 13, IEEE 34, and 14 bus test systems considering several factors, including load imbalance, changes in resistance, utilization of different learning methods, and variation in inception angle. The outcomes of all tested scenarios indicate satisfactory performance, with an error rate below 1%. Moreover, the comparative analysis validates the exceptional performance of the proposed methodology.

REFERENCES

[1] **Ghaderi, A., Ginn, H., and Mohammadpour, H.** (2017). High impedance fault detection: A review. *Electric Power Systems Research, 143*, 376-388.

[2] **Hao, B.** (2020). AI in arcing-HIF detection: A brief review. *IET Smart Grid, 3(4),* 435-444.

[3] **Russell, B. D.** (1982). *Detection of arcing faults on distribution feeders - Final report.* United States: U.S. Department of Energy.

[4] **Jeerings, D. I. and Linders, J. R**. (1989). Ground resistance-revisited. *IEEE Transactions on Power Delivery, 4(2),* 949-956.

[5] **Depew AC, Adamiak MG, Russell BD, Benner CL, Dempsey RW, and Parsick JM.** (2006). Field experience with high-impedance fault detection relays. *2005/2006 IEEE/PES transmission and distribution conference and exhibition,* College Station, TX, USA, April 04-06.

[6] **Power System Relaying Committee** (1989). Downed Power Lines: Why They Can't Always Be Detected. *IEEE Power & Energy Society*.

[7] **Zamanan, N. and Sykulski, J.K.** (2006). Modelling arcing high impedances faults in relation to the physical processes in the electric arc. *Proceedings of the 6th WSEAS International Conference on Power Systems,* Lisbon, Portugal, September 22-24.

[8] **Chen, J. C., Phung, B. T., Zhang, D. M., Blackburn, T., and Ambikairajah, E.** (2013). Study on high impedance fault arcing current characteristics. *2013 Australasian Universities Power Engineering Conference (AUPEC),* Hobart, TAS, Australia, 1-6, 29 September-03 October.

[9] **Michalik, M., Rebizant, W., Lukowicz, M., Lee, S.J., and Kang, S. H.** (2006). High-impedance fault detection in distribution networks with use of wavelet-based algorithm. *IEEE Transactions on Power Delivery, 21(4),* 1793-1802.

[10] **Elkalashy, N. I., Lehtonen, M., Darwish, H. A., Izzularab, M. A., and Taalab, A. I.** (2007). Modeling and experimental verification of high impedance arcing fault in medium voltage networks. *IEEE Transactions on Dielectrics and Electrical Insulation, 14(2),* 375-383.

[11] **Emanuel, A. E., Cyganski, D., Orr, J. A., Shiller, S., and Gulachenski, E. M.** (1990). High impedance fault arcing on sandy soil in 15 kV distribution feeders: contributions to the evaluation of the low frequency spectrum. *IEEE Transactions on Power Delivery, 5(2),* 676-686.

[12] **Nam, S. R., Park, J. K., Kang, Y. C., and Kim, T. H.** (2001). A modeling method of a high impedance fault in a distribution system using two series time-varying resistances in EMTP. *2001 Power Engineering Society Summer Meeting. Conference Proceedings (Cat. No.01CH37262)*, Vancouver, BC, Canada, July 15-19, 1175-1180.

[13] **Lai, T., Snider, L., Lo, E., and Sutanto, D.** (2005). High-impedance fault detection using discrete wavelet transform and frequency range and RMS conversion. *IEEE Transactions on Power Delivery, 20(1),* 397-407.

[14] **Lai, T., Snider, L., and Lo, E.** (2006). Wavelet transform based relay algorithm for the detection of stochastic high impedance faults. *Electric Power Systems Research, 76(8),* 626-633.

[15] **Benner, C., Carswell, P., and Russell, B. D.** (1989). Improved algorithm for detecting arcing faults using random fault behavior. *Electric Power Systems Research, 17(1),* 49-56.

[16] **Santos, W., Souza, B., Brito, N., Costa, F., and Paes, M.** (2013) High Impedance Faults: From Field Tests to Modeling. *Journal of Control, Automation and Electrical Systems, 24(6),* 885-896.

[17] **Lima, E. M., Coelho, R. D., Brito, N. S. D., Souza, B. A.** (2021). High impedance fault detection method for distribution networks under non-linear conditions. *International Journal of Electrical Power & Energy Systems, 131.*

[18] **Cassie, A. M.** (1939). *A new theory ofrupture and circuit severity*. CIGRE Report, 102.

[19] **Mayr, O.** (1943). Beiträge zur Theorie des statischen und des dynamischen Lichtbogens. *Archiv für Elektrotechnik, 37(12),* 588-608.

[20] **Schavemaker P.H. and Slui, L.** (2000). An improved Mayr-type arc model based on current-zero measurements. *IEEE Transactions on Power Delivery, 15,* 580-584.

[21] **Kizilcay M. and Pniok, T**. (1991). Digital Simulation of Fault Arcs in Power systems. *European Transactions on Electrical Power System, 4(3),* 55-59.

[22] **Johns, A., Aggarwal, R., and Song, Y**. (1994). Improved Techniques for Modelling Fault Arcs on Faulted EHV Transmission Systems. *IEE Proceedings-Generation, Transmission and Distribution, 141(2),* 148-154.

[23] **Darwish, H. A. and Elkalashy, N. I.** (2005). Universal Arc Representation Using EMTP. *IEEE Transactions on Power Delivery, 20(2),* 772-779.

[24] **Lee R. E. and Bishop, M. T.** (1985). A Comparison of Measured High Impedance Fault Data to Digital Computer Modeling Results. *IEEE Transactions on Power Apparatus and Systems, 104(10),* 2754-2758.

[25] **Yu D. C. and Khan, S. H.** (1994). An adaptive high and low impedance fault detection method. *IEEE Transactions on Power Delivery, 9(4),* 1812-1821.

[26] **Sharaf, A. M., El-Sharkawy, R. M., Talaat, H. E. A., and Badr, M. A. L.** (1996). Novel alpha-transform distance relaying scheme. *Proceedings of 1996 Canadian Conference on Electrical and Computer Engineering,* Calgary, AB, Canada, May 26-29, 754-757.

[27] **Sharat, A. M., Snider, L. A., and Debnath, K.** (1993). A neural network based back error propagation relay algorithm for distribution system high impedance fault detection. *1993 2nd International Conference on Advances in Power System Control, Operation and Management, APSCOM-93,* Hong Kong, December 07-10.

[28] **Wai, D. C. T. and Yibin, X.** (1998). A novel technique for high impedance fault identification. *IEEE Transactions on Power Delivery, 13(3),* 738-744.

[29] **Lai, T., Snider, L., Lo, E., Cheung, C.H., and Chan, K.W.** (2003). High Impedance Faults Detection Using Artificial Neural Network. *Proceedings of the 6th International Conference on Advances in Power System Control, Operation and Management, APSCOAI,* Hong Kong, November 11-14.

[30] **Sedighi, A. R. and Haghifam, M. R.** (2010). Simulation of high impedance ground fault in electrical power distribution systems. *2010 International Conference on Power System Technology,* Zhejiang, China, October 24-28.

[31] **Zamanan, N. and Sykulski, J.** (2014). The evolution of high impedance fault modeling. *16th International Conference on Harmonics and Quality of Power (ICHQP),* Bucharest, Romania, May 25-28.

[32] **Michalik, M., Rebizant, W., Lukowicz, M., Lee, S., and Kang, S.** (2005). Wavelet transform approach to high impedance fault detection in MV networks. *IEEE Russia Power Tech,* 1-7.

[33] **Cui, T., Dong, X., Bo, Z., Klimek, A., and Edwards, A.** (2008). Modeling study for high impedance fault detection in MV distribution system. *43rd International Universities Power Engineering Conference,* Padova, September 01-04.

[34] **Sedighizadeh, M., Rezazadeh A., and Elkalashy, N.** (2010). Approaches in High Impedance Fault Detection - A Chronological Review. *Advances in Electrical and Computer Engineering, 10(3),* 114-128.

[35] **Vyshnavi, G. and Prasad, A.** (2018). Detection and Location of High Impedance Faults in Distribution Systems: A Review. *International Journal of Advanced Science and Technology, 119,* 53-66.

[36] **Carr, J.** (1981). Detection of High Impedance Faults on Multi-Grounded Primary Distribution Systems. *IEEE Power Engineering Review, 1(4),* 69-70.

[37] **Calhoun, H., Bishop, M. T., Eichler, C. H., and Lee, R. E.** (1982). Development and Testing of an Electro-Mechanical Relay to Detect Fallen Distribution Conductors. *IEEE Transactions on Power Apparatus and Systems, 101(6),* 1643-1650.

[38] **Lee, R. E. and Bishop, M. T.** (1983). Performance Testing of the Ratio Ground Relay on a Four-Wire Distribution Feeder. *IEEE Power Engineering Review, 3(9),* 28-29.

[39] **Huang, C., Chu H., and Chen M. T. (1988).** Algorithm comparison for high impedance fault detection based on staged fault test. *IEEE Transactions on Power Delivery, 3(4),* 1427-1435.

[40] **Reedy, R. M. and Elsmore, W. A.** (1983). Electromechanical relay to detect fallen distribution conductors. *American Public Power Association Conference,* San Antonio, February.

[41] **Wester, C. G.** (1998). High impedance fault detection on distribution systems. *1998 Rural Electric Power Conference Presented at 42nd Annual Conference,* St. Louis, MO, USA, April 26-28.

[42] **Wang, B., Geng, J., and Dong, X.** (2018). High-Impedance Fault Detection Based on Nonlinear Voltage–Current Characteristic Profile Identification. *IEEE Transactions on Smart Grid, 9(4),* 3783-3791.

[43] **Theron, J. C. J., Pal, A., and Varghese, A.** (2018). Tutorial on high impedance fault detection. *71st Annual Conference for Protective Relay Engineers (CPRE),* College Station, TX, USA, March 26-29.

[44] **Novak, T., Morley L. A., and Trutt, F. C.** (1988). Sensitive ground-fault relaying. *IEEE Transactions on Industry Applications, 24(5),* 853-861.

[45] **Aljohani A. and Habiballah, I.** (2020). High-Impedance Fault Diagnosis: A Review. *Energies, 13(23),* 6447.

[46] **Aucoin, B. and Russell, B.** (1982). Distribution High Impedance Fault Detection Utilizing High Frequency Current Components. *IEEE Transactions on Power Apparatus and Systems, 101(6),* 1596-1606.

[47] **Lee, I.** (1982). *High impedance fault detection using third harmonic current final report.* Electric Power Research Institute, The Institute, Palo Alto, Calif., USA.

[48] **Russell, B., Chinchali, R. P., and Kim, C. J.** (1988). Behaviour of low frequency spectra during arcing fault and switching events. *IEEE Transactions on Power Delivery, 3(4),* 1485-1492.

[49] **Mai, W., Phung, B. T., and Ambikairajah, E.** (2012). Detection of high impedance faults in medium voltage distribution networks. *10th International Power & Energy Conference (IPEC),* Ho Chi Minh City, Vietnam, December 12-14.

[50] **Yeh, H., Tran, D. H., and Yinger, R.** (2014). High impedance fault detection using orthogonal transforms. *2014 IEEE Green Energy and Systems Conference (IGESC),* Long Beach, CA, November 24.

[51] **Soheili, A., Sadeh, J., Lomei, H., and Muttaqi, K.** (2016). A new high impedance fault detection scheme: Fourier based approach. *2016 IEEE International Conference on Power System Technology (POWERCON),* Wollongong, NSW, Australia, September 28- October 01.

[52] **Lima, M., Brito, N. S. D., Souza, B. A., Santos, W. C., and Fortunato, L. M. A.** (2016). Analysis of the influence of the window used in the Short-Time Fourier Transform for High Impedance Fault detection. *17th International Conference on Harmonics and Quality of Power (ICHQP),* Belo Horizonte, Brazil, October 16-19.

[53] **Wali, M. K., Hussain, A. N., and Hani, W. F.** (2017). High impedance fault detection based on power spectrum technique. *2017 International Conference on Engineering and Technology (ICET),* Antalya, Turkey, August 21-23.

[54] **Zhang, B., Sun, Y., Shi, F., Zhang, H., Liu, S., and Zhang, Y.** (2018). Detection of arc grounding fault in distribution network based on the harmonic component. *13th IEEE Conference on Industrial Electronics and Applications (ICIEA),* Wuhan, China, May 31- June 02.

[55] **Soheili, A., Sadeh, J., and Bakhshi, R.** (2018). Modified FFT based high impedance fault detection technique considering distribution non-linear loads: Simulation and experimental data analysis. *International Journal of Electrical Power & Energy Systems, 94,* 124-140.

[56] **Graps, A.** (1995). An introduction to wavelets. *IEEE Computational Science and Engineering, 2(2),* 50-61.

[57] **Huang, S., Hsieh, C., and Huang, C.** (1998). Application of wavelets to classify power system disturbances. *Electric Power Systems Research, 47(2),* 87-93.

[58] **Lazkano, A., Ruiz, J., Aramendi, E., and Leturiondo, L. A.** (2000). A new approach to high impedance fault detection using wavelet packet analysis. *Ninth International Conference on Harmonics and Quality of Power. Proceedings (Cat. No.00EX441),* Orlando, FL, USA, October 01-04.

[59] **Elkalashy, N. I., Lehtonen, M., Darwish, H. A., Izzularab, M. A., and Taalab, A. I.** (2007). DWT-Based Investigation of phase currents for Detecting High Impedance Faults Due to Leaning Trees in Unearthed MV Networks. *2007 IEEE Power Engineering Society General Meeting,* Tampa, FL, USA, June 24-28.

[60] **Gilany, M., Zamanan N., and Wahba, W.** (2010). A wavelet based technique for detection and classification of transients in distribution networks. *The 2nd International Conference on Computer and Automation Engineering (ICCAE),* Singapore, February 26-28.

[61] **Garcia, V. T. and Paredes, H. R.** (2011). High impedance fault detection in Distribution System using Wavelet Transform. *8th International Conference on Electrical Engineering, Computing Science and Automatic Control,* Merida City, Mexico, October 26-28.

[62] **Chen, J. C., Phung, B. T., Wu, H. W., Zhang, D. M., and Blackburn, T.** (2014). Detection of High Impedance Faults using wavelet transform. *2014 Australasian Universities Power Engineering Conference (AUPEC),* Perth, WA, Australia, September 28-October 01.

[63] **Ramamurthy, T. A. and Swarup, K. S.** (2016). High Impedance Fault detection using DWT for transmission and distribution networks. *IEEE 6th International Conference on Power Systems (ICPS),* New Delhi, India, March 04-06.

[64] **Santos, W. C., Lopes, F. V., Brito, N. S. D., and Souza, B. A.** (2017). High-Impedance Fault Identification on Distribution Networks. *IEEE Transactions on Power Delivery, 32(1),* 23-32.

[65] **Ferreira, G. A. and Assis, T. M. L.** (2019). A Novel High Impedance Arcing Fault Detection Based on the Discrete Wavelet Transform for Smart Distribution Grids. *2019 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America),* Gramado, Brazil, September 15-18.

[66] **Lai, L. L., Styvaktakis, E., and Sichanie, A. G.** (1997). Wavelet transform for high impedance fault identification. *Fourth International Conference on Advances in Power System Control, Operation and Management, APSCOM-97. (Conf. Publ. No. 450),* Hong Kong, November 11-14.

[67] **Huang, S. J. and Hsieh, C. T.** (1999). High-impedance fault detection utilizing a Morlet wavelet transform approach. *IEEE Transactions on Power Delivery, 14(4),* 1401-1410.

[68] **Michalik, M., Lukowicz, M., Rebizant, W., Lee, S. and Kang, S.** (2007). Verification of the Wavelet-Based HIF Detecting Algorithm Performance in Solidly Grounded MV Networks. *IEEE Transactions on Power Delivery, 22(4),* 2057-2064.

[69] **Linders, J. R. U. S. and Jeerings, D. I. U. S.** (1989). High impedance fault analyzer in electric power distribution. *U.S. Patent No. 4871971A.*

[70] **Hou, D.** (2007). Detection of high-impedance faults in power distribution systems. *2007 Power Systems Conference: Advanced Metering, Protection, Control, Communication, and Distributed Resources,* Clemson, SC, USA, March 13-16.

[71] **Sheng, Y. and Rovnyak, S. M.** (2004). Decision tree-based methodology for high impedance fault detection. *IEEE Transactions on Power Delivery, 19(2),* 533-536.

[72] **Shahrtash, S. M. and Sarlak, M.** (2006). High Impedance Fault Detection Using Harmonics Energy Decision Tree Algorithm. *2006 International Conference on Power System Technology,* Chongqing, China, October 22-26.

[73] **Samantaray, S.** (2012). Ensemble decision trees for high impedance fault detection in power distribution network. *International Journal of Electrical Power & Energy Systems, 43(1),* 1048-1055.

[74] **Kar, S. and Samantaray, S. R.** (2016). High impedance fault detection in microgrid using maximal overlapping discrete wavelet transform and decision tree. *2016 International Conference on Electrical Power and Energy Systems (ICEPES),* Bhopal, India, December 14-16.

[75] **Sekar, K. Mohanty, N. K., and Sahoo, A. K.** (2018). High impedance fault detection using wavelet transform. *2018 Technologies for Smart-City Energy Security and Power (ICSESP),* Bhubaneswar, India, March 28-30.

[76] **Sedighi, A., Haghifam, M., and Malik, O.** (2005). Soft computing applications in high impedance fault detection in distribution systems. *Electric Power Systems Research, 76,* 136-144.

[77] **Haghifam, M., Sedighi, A., and Malik, O.** (2006). Development of a fuzzy inference system based on genetic algorithm for high-impedance fault detection. *IEE Proceedings - Generation, Transmission and Distribution,* 153(3), 359.

[78] **Zamanan, N., Sykulski, J. K., and Al-Othman, A. K**. (2007). Arcing High Impedance Fault Detection Using Real Coded Genetic Algorithm. *Third IASTED Asian Conference Power and Energy Systems,* Phuket, Thailand. April 01-03.

[79] **Yang, M., Gu, J., Jeng, C., and Kao, W.** (2004). Detection of high impedance fault in distribution feeder using wavelet transform and artificial neural networks. *2004 International Conference on Power System Technology, PowerCon,* Singapore, November 21-24.

[80] **Mokhtari, H. and Aghatehrani, R.** (2005). A new wavelet-based method for detection of high impedance faults. *2005 International Conference on Future Power Systems,* Amsterdam, Netherlands, November 18.

[81] **Baqui, I., Mazón, A. J., Zamora, I., and Vicente, R.** (2005). High impedance faults detection in power distribution system by combination of artificial neural network and wavelet transform. *CIRED 2005 - 18th International Conference and Exhibition on Electricity Distribution,* Turin, Italy, June 06-09.

[82] **Ma, S. and Guan, L.** (2011). Arc-Fault Recognition Based on BP Neural Network. *2011 Third International Conference on Measuring Technology and Mechatronics Automation,* Shanghai, China, January 06-07.

[83] **Abohagar, A. A. and Mustafa, M. W.** (2012). Back propagation neural network aided wavelet transform for high impedance fault detection and faulty phase selection. *2012 IEEE International Conference on Power and Energy (PECon),* Kota Kinabalu, Malaysia, December 02-05.

[84] **Vijayachandran, G. and Mathew, B. K.** (2012). High impedance arcing fault detection in MV networks using discrete wavelet transform and Artificial Neural Networks. *2012 International Conference on Green Technologies (ICGT),* Trivandrum, India, December 18-20.

[85] **Lai, T. M., Lo, W. C., To, W., and Lam, K. H**. (2012). RMS percent of wavelet transform for the detection of stochastic high impedance faults. *2012 IEEE 15th International Conference on Harmonics and Quality of Power,* Hong Kong, China, June 17-20.

[86] **Guo, M., Zeng, X., Chen, D., and Yang, N.** (2018). Deep-Learning-Based Earth Fault Detection Using Continuous Wavelet Transform and Convolutional Neural Network in Resonant Grounding Distribution Systems. *IEEE Sensors Journal, 18(3),* 1291-1300.

[87] **Sirojan, T., Lu, S., Phung, B.T., Zhang, D., and Ambikairajah, E.** (2018). High Impedance Fault Detection by Convolutional Deep Neural

Network. *2018 IEEE International Conference on High Voltage Engineering and Application (ICHVE),* Athens, Greece, September 10-13.

[88] **Wang, S. and Dehghanian, P.** (2020). On the Use of Artificial Intelligence for High Impedance Fault Detection and Electrical Safety. *IEEE Transactions on Industry Applications, 56(6),* 7208-7216.

[89] **Gu, J. C., Huang, Z. J., Wang, J. M., Hsu, L. C., and Yang, M. T.** (2021). High Impedance Fault Detection in Overhead Distribution Feeders Using a DSP-Based Feeder Terminal Unit. *IEEE Transactions on Industry Applications, 57(1),* 179-186.

[90] **Eissa, M. M., Sowilam, G. M., and Sharaf, A. M.** (2006). A New Protection Detection Technique for High Impedance Fault Using Neural Network. *2006 Large Engineering Systems Conference on Power Engineering,* Halifax, NS, Canada, July 26-28.

[91] **Garcia, J. C., Vega Garcia, V., and Kagan, N.** (2014). Detection of high impedance faults in overhead multi grounded networks. *2014 11th IEEE/IAS International Conference on Industry Applications,* Juiz de Fora, Brazil, December 07-10.

[92] **Souza, J. V., Lopes, G. N., Vieira, J. C., and Asada, E. N**. (2020). High Impedance Fault Detection in Distribution Systems: An Approach Based on Fourier Transform and Artificial Neural Networks. *2020 Workshop on Communication Networks and Power Systems (WCNPS),* Brasilia, Brazil, November 12-13.

[93] **Sarlak, M., and Shahrtash, S. M**. (2008). High impedance fault detection in distribution networks using support vector machines based on wavelet transform. *2008 IEEE Canada Electric Power Conference,* Vancouver, BC, Canada, October 06-07.

[94] **Moloi, K., Jordaan, J. A., and Hamam, Y.** (2017). High impedance fault detection technique based on Discrete Wavelet Transform and support vector machine in power distribution networks. *2017 IEEE AFRICON,* 9-14, Cape Town, South Africa, September 18-20.

[95] **Mortazavi, S., Moravej, Z., and Shahrtash, S.** (2018). A hybrid method for arcing faults detection in large distribution networks. *International Journal of Electrical Power & Energy Systems, 94,* 141-150.

[96] **Du, J., Liu, Y., and Xu, Z.** (2020). Research on Single-Phase High Impedance Fault Identification of Neutral Ungrounded System. *2020 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia),* Weihai, China, 13-15 July.

[97] **Elkalashy, N. I., Lehtonen, M., Darwish, H. A., Taalab, A. I., and Izzularab, M. A.** (2008). DWT-Based Detection and Transient Power Direction-Based Location of High-Impedance Faults Due to Leaning Trees in Unearthed MV Networks. *IEEE Transactions on Power Delivery, 23(1),* 94-101.

[98] **Ali, M. S., Bakar, A. H., Mokhlis, H. B., Aroff, H., Illias, H. A., and Aman, M. M.** (2012). High impedance fault localization in a distribution

network using the discrete wavelet transform. *2012 IEEE International Power Engineering and Optimization Conference* Melaka, Malaysia, June 06-07.

[99] **Ali, M., Abu Bakar, A., Tan, C., Mokhlis, H., Abu Talip, M., and Arof, H.** (2016). Studies on the application of wavelet families for a high impedance fault location algorithm in a distribution network. *Turkish Journal of Electrical Engineering and Computer Sciences, 24,* 5043-5054.

[100] **Santos, W. C., Lopes, F.V., Brito, N. S., Souza, B. A., Fernandes, D., and Neves, W. L.** (2013). High Impedance Fault Detection and Location Based on Electromagnetic Transient Analysis. *2013 International Conference on Power Systems Transients (IPST2013)* Vancouver, Canada, July 18-20.

[101] **Mahari A. and Seyedi, H.** (2015). High impedance fault protection in transmission lines using a WPT-based algorithm. *International Journal of Electrical Power & Energy Systems, 67,* 537-545.

[102] **Adly, A., Aleem, S., Algabalawy, M., Jurado, F., and Ali, Z.** (2020). A novel protection scheme for multi-terminal transmission lines based on wavelet transform. *Electric Power Systems Research, 183.*

[103] **Hossain, S., Zhu, H., and Overbye, T. J.** (2014). Distribution high impedance fault location using localized voltage magnitude measurements. *2014 North American Power Symposium (NAPS),* Pullman, WA, USA, September 07-09.

[104] **Garcia-Santander, L., Bastard, P., Petit, M., Gal, I., Lopez, E., and Opazo, H.** (2005). Down-conductor fault detection and location via a voltage based method for radial distribution networks. *IEE Proceedings - Generation, Transmission and Distribution, 152(2),* 180-184.

[105] **Vieira, F. L., Filho, J. M., Silveira, P. M., Guerrero, C. A., and Leite, M. P.** (2018). High impedance fault detection and location in distribution networks using smart meters. *18th International Conference on Harmonics and Quality of Power (ICHQP),* Ljubljana, Slovenia, May 13-16.

[106] **Bhandia, R., Chavez, J. d. J., Cvetković, M., and Palensky, P.** (2020). High Impedance Fault Detection Using Advanced Distortion Detection Technique. *IEEE Transactions on Power Delivery, 35(6),* 2598-2611.

[107] **Uriarte, F.M., and Centeno, V.** (2005). High-impedance fault detection and localization in distribution feeders with microprocessor based devices. *Proceedings of the 37th Annual North American Power Symposium,* Ames, IA, USA, October 25.

[108] **Radhakrishnan, A., and Das, S.** (2019). Location of High Impedance Faults using Smart Meters in Distribution Systems with DGs, Power Electronic Loads and Electric Arc Furnaces. *2019 IEEE Milan PowerTech,* Milan, Italy, June 23-27.

[109] **Hong Y. and Huang, W.** (2015). Locating High-Impedance Fault Section in Electric Power Systems Using Wavelet Transform, k-Means, Genetic

Algorithms, and Support Vector Machine. *Mathematical Problems in Engineering, 2015*, 1-9.

[110] **Hong, Y., Huang, W., Chang, Y., Lee, Y. D., and Ouyang, D.** (2017). Locating high-impedance fault in a smart distribution system using wavelet entropy and hybrid self-organizing mapping network. *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, Turin, Italy, September 26-29.

[111] **Mahmoud, M.M.** (2015). Detection of high impedance faults in M.V. mesh distribution network. *2015 Modern Electric Power Systems (MEPS)*, Wroclaw, Poland, July 06-09.

[112] **Oliveira, B. C., Pereira, J. L., Melo, I. D., Souza, M. A., and Alves, O. G.** (2018). A new methodology for high impedance fault detection, classification and location using PMUs. *2018 Simposio Brasileiro de Sistemas Eletricos (SBSE)*, Niteroi, Brazil, May 12-16.

[113] **Vianna, J., Araujo, L., and Penido, D.** (2016). High Impedance Fault Area Location in Distribution Systems Based on Current Zero Sequence Component. *IEEE Latin America Transactions, 14(2)*, 759-766.

[114] **Vianna, J., Guaracy, P., Araujo L., and Penido, D.** (2021). A method to detect and locate faulted area in distribution systems using the existing measurements structure. *International Journal of Electrical Power & Energy Systems, 131*.

[115] **Wei, M., Zhang, H., Shi, F., Chen, W., and Terzija, V.** (2022). Nonlinearity Characteristic of High Impedance Fault at Resonant Distribution Networks: Theoretical Basis to Identify the Faulty Feeder. *IEEE Transactions on Power Delivery, 37(2)*, 923-936.

[116] **Wang, X. Liu, W., Liang, Z., Guo, L., Du, H., Gao, J., and Li, C.** (2022). Faulty feeder detection based on the integrated inner product under high impedance fault for small resistance to ground systems. *International Journal of Electrical Power & Energy Systems, 140*.

[117] **Wu, W., Zhang, P., Qiao, D., Sun, Q., and Wang, W.** (2022). A Faulty Feeder Selection Method based on Improved Hausdorff Distance Algorithm for Neutral Non-effectively Grounded System. *Electric Power Systems Research, 203.*

[118] **Li, J., Wang, G., Zeng, D., and Li, H.** (2020). High-impedance ground faulted line-section location method for a resonant grounding system based on the zero-sequence current's declining periodic component. *International Journal of Electrical Power & Energy Systems, 119.*

[119] **Xue, S., Cheng, X., Zong, X., and Lv, Y.** (2017). Research on high-impedance fault location based on complex sequence current. *2017 Chinese Automation Congress (CAC)*, Jinan, China, October 20-22.

[120] **Xue, S., Cheng, X., and Lv, Y.** (2017). High Resistance Fault Location of Distribution Network Based on EEMD. *2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2017, Hangzhou, China, August 26-27.

[121] **Wei, M., Liu, W., Shi, F., Zhang, H., Jin, Z., and Chen, W.** (2021). Distortion-Controllable arc modeling for high impedance arc fault in the distribution network. *IEEE Transactions on Power Delivery, 36(1),* 52-63.

[122] **Radojevic, Z. M., Terzija, V. V., and Djuric, N. B.** (2000). Numerical algorithm for overhead lines arcing faults detection and distance and directional protection. *IEEE Transactions on Power Delivery, 15(1),* 31-37.

[123] **Ferraz, R. G., Iurinic, L. U., Filomena, A. D., and Bretas, A. S.** (2014). High impedance fault location formulation: A least square estimator based approach. *12th IET International Conference on Developments in Power System Protection (DPSP 2014),* Copenhagen, Denmark, March 31-April 03.

[124] **Ferraz, R., Iurinic, L., Filomena, A., Gazzana, D., and Bretas, A.** (2016). Arc fault location: A nonlinear time varying fault model and frequency domain parameter estimation approach. *International Journal of Electrical Power & Energy Systems, 80,* 347-355.

[125] **Ramos, M., Bretas, A., Bernardon, D., and Pfitscher, L.** (2017). Distribution networks HIF location: A frequency domain system model and WLS parameter estimation approach. *Electric Power Systems Research, 146,* 170-176.

[126] **Ramos, M., Resener, M., Bretas, A., Bernardon, D., and Leborgne, R. C.** (2020). Physics-based analytical model for high impedance fault location in distribution networks. *Electric Power Systems Research, 188,* 106577.

[127] **Yin, Z., Wei, Z., Sun, G., Zang, H., and Chen, S.** (2022). High sensitivity fault location technology for distribution networks considering measurement error. *International Journal of Electrical Power & Energy Systems, 140*.

[128] **Doria-García, J., Orozco-Henao, C., Leborgne, R., Montoya, O., and Gil-González, W.** (2021). High impedance fault modeling and location for transmission line☆. *Electric Power Systems Research, 196*.

[129] **Nunes, J. U. N., Bretas, A. S., Bretas, N. G., Herrera-Orozco, A. R., and Iurinic, L. U.** (2019). Distribution systems high impedance fault location: A spectral domain model considering parametric error processing. *International Journal of Electrical Power & Energy Systems, 109*.

[130] **Dobakhshari, A. S. and Ranjbar, A. M.** (2015). A Novel Method for Fault Location of Transmission Lines by Wide-Area Voltage Measurements Considering Measurement Errors. *IEEE Transactions on Smart Grid, 6(2),* 874-884.

[131] **Ferreira, G. D., Gazzana, D. D., Bretas, A. S., and Netto, A. S.** (2012). A unified impedance-based fault location method for generalized distribution systems. *2012 IEEE Power and Energy Society General Meeting,* San Diego, CA, USA, July 22-26.

[132] **Dasco, A., Marguet, R., and Raison, B**. (2015). Fault distance estimation in distribution network for high impedance faults. *2015 IEEE Eindhoven PowerTech,* Eindhoven, Netherlands, June 29- July 02.

[133] **Iurinic, L. U., Herrera-Orozco, A. R., Ferraz, R. G., and Bretas, A. S.** (2016). Distribution Systems High-Impedance Fault Location: A Parameter Estimation Approach. *IEEE Transactions on Power Delivery, 31(4),* 1806-1814.

[134] **Farajollahi, M., Shahsavari, A., and Rad, H. M.** (2017). Location identification of high impedance faults using synchronized harmonic phasors. *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT),* Washington, DC, USA, April 23-26.

[135] **Mortazavi, S. H., Moravej, Z., and Shahrtash, S. M.** (2019). A Searching Based Method for Locating High Impedance Arcing Fault in Distribution Networks. *IEEE Transactions on Power Delivery, 34(2),* 438-447.

[136] **Gazzana D. S., Ferreira, G.D., Bretas, A.S., Bettiol, A.L., Carniato, A., Passos, L.F.N., Ferreira, A.H., and Silva, J.E.M.** (2014). An integrated technique for fault location and section identification in distribution systems. *Electric Power Systems Research, 115,* 65-73.

[137] **Chatterjee, B. and Debnath, S.** (2020). Sequence component based approach for fault discrimination and fault location estimation in UPFC compensated transmission line. *Electric Power Systems Research, 180.*

[138] **Chatterjee, B. and Debnath, S.** (2021). A new protection scheme for transmission lines utilizing positive sequence fault components. *Electric Power Systems Research, 190,* 106847.

[139] **Glik, K., Kowalik, R., Rasolomampionona, D., and Anwar, S**. (2011). Travelling wave fault location in power transmission systems: an overview. *Journal of Electrical Systems, 7,* 287-296.

[140] **Dalcastagnê, A. L. and Zimath, S. L.** (2008). A study about the sources of error of impedance-based fault location methods. *2008 IEEE/PES Transmission and Distribution Conference and Exposition: Latin America,* Bogota, Colombia, August 13-15.

[141] **Schweitzer, E. O., Guzman, A., Mynam, M., Skendzic, V., Kasztenny, B., and Marx, S.** (2014). A new traveling wave fault locating algorithm for line current differential relays. *12th IET International Conference on Developments in Power System Protection (DPSP 2014),* Copenhagen, Denmark, March 31-April 03.

[142] **Elkalashy, N. I., Sabiha, N. A., and Lehtonen, M. (**2015). Earth Fault Distance Estimation Using Active Traveling Waves in Energized-Compensated MV Networks. *IEEE Transactions on Power Delivery, 30(2),* 836-843.

[143] **Bernadić, A. and Leonowicz, Z.** (2012). Fault location in power networks with mixed feeders using the complex space-phasor and Hilbert–Huang transform. *International Journal of Electrical Power & Energy Systems, 42(1),* 208-219.

[144] **Bashir, M., Niazy, I., Sadeh, J., and Taghizadeh, M.** (2011). Considering characteristics of arc on travelling wave fault location algorithm for the transmission lines without using line parameters. *2011 10th International Conference on Environment and Electrical Engineering,* Rome, Italy, May 08-11.

[145] **Glik, K., and Rasolomampionona, D. D.** (2013). Travelling wave fault location algorithm in HV lines — Simulation test results for arc and high impedance faults. *Eurocon 2013,* Zagreb, Croatia, 01-04 July.

[146] **Rezaei, D., Gholipour M., and Parvaresh, F.** (2022). A single-ended traveling-wave-based fault location for a hybrid transmission line using detected arrival times and TW′s polarity. *Electric Power Systems Research,* 210.

[147] **Zimath, S. L., Dutra, C. A., Matos, R. R., Oliveira, L. B., Resende, J., and Moutinho, J. A.** (2014). Traveling wave fault location applied to high impedance events. *12th IET International Conference on Developments in Power System Protection (DPSP 2014),* Copenhagen, Denmark, March 31-April 03.

[148] **Silva, J. A., Costa, F. B., Santos, W. C., Neves, W. L., and Souza, B. A.** (2013). High impedance fault location — Case study using wavelet transform and artificial neural networks. *22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013),* Stockholm, June 10-13.

[149] **Lucas, F. P., Costa, P. P., Batalha, R. M., and Leite, D. F.** (2018). High Impedance Fault Detection in Time-Varying Distributed Generation Systems Using Adaptive Neural Networks. *2018 International Joint Conference on Neural Networks (IJCNN),* Rio de Janeiro, Brazil, July 08-13.

[150] **Ledesma, J. J. G., Nascimento, K. B., Araujo, L. R., and Penido, D. R. S.** (2020). A two-level ANN-based method using synchronized measurements to locate high-impedance fault in distribution systems. *Electric Power Systems Research, 188,* 106576.

[151] **Yadav, A. and Swetapadma, A.** (2015). A single ended directional fault section identifier and fault locator for double circuit transmission lines using combined wavelet and ANN approach. *International Journal of Electrical Power & Energy Systems, 69,* 27-33.

[152] **Jamali, S. and Ghaffarzadeh, N.** (2011). A new method for arcing fault location using discrete wavelet transform and wavelet networks. *European Transactions on Electrical Power, 22(5),* 601-615.

[153] **Bretas, A. S., Moreto, M., Salim, R. H., and Pires, L.** (2006). A Novel High Impedance Fault Location for Distribution Systems Considering Distributed Generation. *2006 IEEE/PES Transmission & Distribution Conference and Exposition: Latin America,* Caracas, Venezuela, August 15-18.

[154] **Farias, P., Morais, A., Rossini, J., and Cardoso, G.** (2018). Nonlinear high impedance fault distance estimation in power distribution systems: A continually online-trained neural network approach. *Electric Power Systems Research, 157,* 20-28.

[155] **Osman, A. H., Abdelazim, T., and Malik, O. P.** (2005). Transmission line distance relaying using on-line trained neural networks. *IEEE Transactions on Power Delivery, 20(2),* 1257-1264.

[156] **Silva, J. A., Costa, F. B., and Santos, W. C.** (2011). High impedance fault location — case study with developed models from field experiments. *21st International Conference on Electricity Distribution,* Frankfurt, Germany, June 6-9.

[157] **Bhatnagar, M., Yadav, A., and Swetapadma, A.** (2022). Enhancing the resiliency of transmission lines using extreme gradient boosting against faults. *Electric Power Systems Research, 207*.

[158] **Bouricha, A., Bouthiba, T., Boukhari, R., and Seghir, S.** (2018). High Impedance Faults Location in the Distribution Networks using Adaptive Neuro-Fuzzy Inference System. 2018 *International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM),* Algiers, Algeria, October 28-31.

[159] **Batista, O., Flauzino, R., Araujo, M., Moraes, L., and Silva, I.** (2016). Methodology for information extraction from oscillograms and its application for high-impedance faults analysis. *International Journal of Electrical Power & Energy Systems, 76,* 23-34.

[160] **Moloi, K., Jordaan, J. A., and Hamam, Y**. (2018). High Impedance Fault Classification and Localization Method for Power Distribution Network. *2018 IEEE PES/IAS PowerAfrica,* Cape Town, South Africa, June 28-29.

[161] **Bhongade, S. and Golhani, S.** (2016). HIF detection using wavelet transform, travelling wave and support vector machine. *2016 International Conference on Electrical Power and Energy Systems (ICEPES),* Bhopal, India, December 14-16.

[162] **Livani H. and Evrenosoglu, C. Y.** (2014). A Machine Learning and Wavelet-Based Fault Location Method for Hybrid Transmission Lines. *IEEE Transactions on Smart Grid, 5(1),* 51-59.

[163] **Raisz, D. and Gonczi, J.** (2014). Fault location methods at compensated MV networks. *49th International Universities Power Engineering Conference (UPEC),* Cluj-Napoca, Romania, September 02-05.

[164] **Haykin S.** (2009). *Neural Networks and Learning Machines.* Prentice Hall.

[165] **Rumelhart, D. E., and McClelland, J. L.** (1986). *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations.* Cambridge (Mass.): MIT Press.

[166] **Muhammad, S., Faisal, M., Muhammad, A., Abdul, K. Q., Sufi, G. T., Adil, K. S.** (2020). High impedance fault detection and isolation in power distribution networks using support vector machines. *Journal of King Saud University - Engineering Sciences, 32(8),* 524-535.

[167] **Fix, E. and Hodges, J.L.** (1989). Discriminatory Analysis - Nonparametric Discrimination: Consistency Properties. *International Statistical Review, 57(3),* 238-247.

[168] **Daubechies, I**. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics, 41(7),* 909-996.

[169] **Piesciorovsky, E. C., Piesciorovsky, E. C., Schulz, N. N., and Schulz, N. N.** (2017). Fuse relay adaptive overcurrent protection scheme for microgrid with distributed generators. *Iet Generation Transmission & Distribution, 11,* 540-549.

[170] **Xie, W., Wang, X., Fang, C., Zhang, H., Shi, F., Xing, X., and Sun, B.** (2020). Field experiment using transient energy method to locate a single-phase to ground fault. *Global Energy Interconnection, 3(6),* 585-594.

[171] **Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., and Liu, H.** (2017). Feature Selection: A Data Perspective. *ACM Computing Surveys, 50(6),* 1–45.

[172] **Robnik-Sikonja, M., and Kononenko, I.** (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning, 53,* 23-69.

[173] **Ho, T. K.** (1995). Random decision forests, *Proceedings of 3rd International Conference on Document Analysis and Recognition.* Montreal, QC, Canada, August 14-16.

[174] **Breiman L.** (2001). Random forests. *Machine Learning, 45(1),* 5-32.

[175] **Geurts P., Ernst D., and Wehenkel L.,** (2006). Extremely randomized trees. *Machine Learning, 63(1),* 3-42.

[176] **Tibshirani, R.** (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the royal statistical society series b-methodological, 58,* 267-288.

[177] **Milioudis, A. N., Andreou, G. T., and Labridis, D. P.** (2015). Detection and Location of High Impedance Faults in Multiconductor Overhead Distribution Lines Using Power Line Communication Devices. *IEEE Transactions on Smart Grid, 6(2),* 894-902.

[178] **Tomasso, A., Invernizzi, G., and Vielmini, G.** (2019). Accurate Single-End and Double-End Fault Location by Traveling Waves: a review with some real applications. *2019 AEIT International Annual Conference (AEIT),* Florence, Italy, September 18-20.

[179] **Fedorov, A., Petrov, V., and Naumov, V.** (2022). The Rules for Using Modal Transformation in Traveling Wave Fault Locator. *2022 International Ural Conference on Electrical Power Engineering (UralCon),* Magnitogorsk, Russian Federation, September 23-25.

[180] **Alekseev, V. I., Petrov, V., and Naumov, V.** (2020). Invariance of Modal Transformations of Electrical Values in Traveling Wave Fault Locator. *2020 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM),* Sochi, Russia, May 18-22.

[181] **Clarke, E.** (1943). *Circuit analysis of AC power systems: symmetrical and related components.* New York, Wiley.

[182] **Sarwagya, K., De, S., and Nayak, P. K.** (2018). High-impedance fault detection in electrical power distribution systems using moving sum approach. *IET Science, Measurement & Technology, 12(1),* 1-8.

[183] **Mora-Flòrez, J., Meléndez, J., and Carrillo-Caicedo, G.** (2008). Comparison of impedance based fault location methods for power distribution systems. *Electric Power Systems Research, 78(4),* 657-666.

[184] **Kakolaki, S. E. H., Hakimian, V., Sadeh, J., and Rakhshani, E.** (2023). Comprehensive Study on Transformer Fault Detection via Frequency Response Analysis. *IEEE Access, 11*, 81852-81881.

[185] **Farias, P. E., Morais, A. P., Rossini, J. P., and Cardoso, G.** (2018). Non-linear high impedance fault distance estimation in power distribution systems: A continually online-trained neural network approach. *Electric Power Systems Research, 157*, 20–28.

APPENDICES

APPENDIX A: Modeled Networks

APPENDIX B: Verification of Modeled Networks

APPENDIX A: Modeled Networks

The simulation studies were conducted on the IEEE 13, IEEE 34, and 14 bus test systems. HIFs were simulated at various locations within the modeled networks. All modeling and analysis were carried out in the Simulink environment. The modeled networks are illustrated below in Figure A.1, Figure A.2 and Figure A.3, respectively.

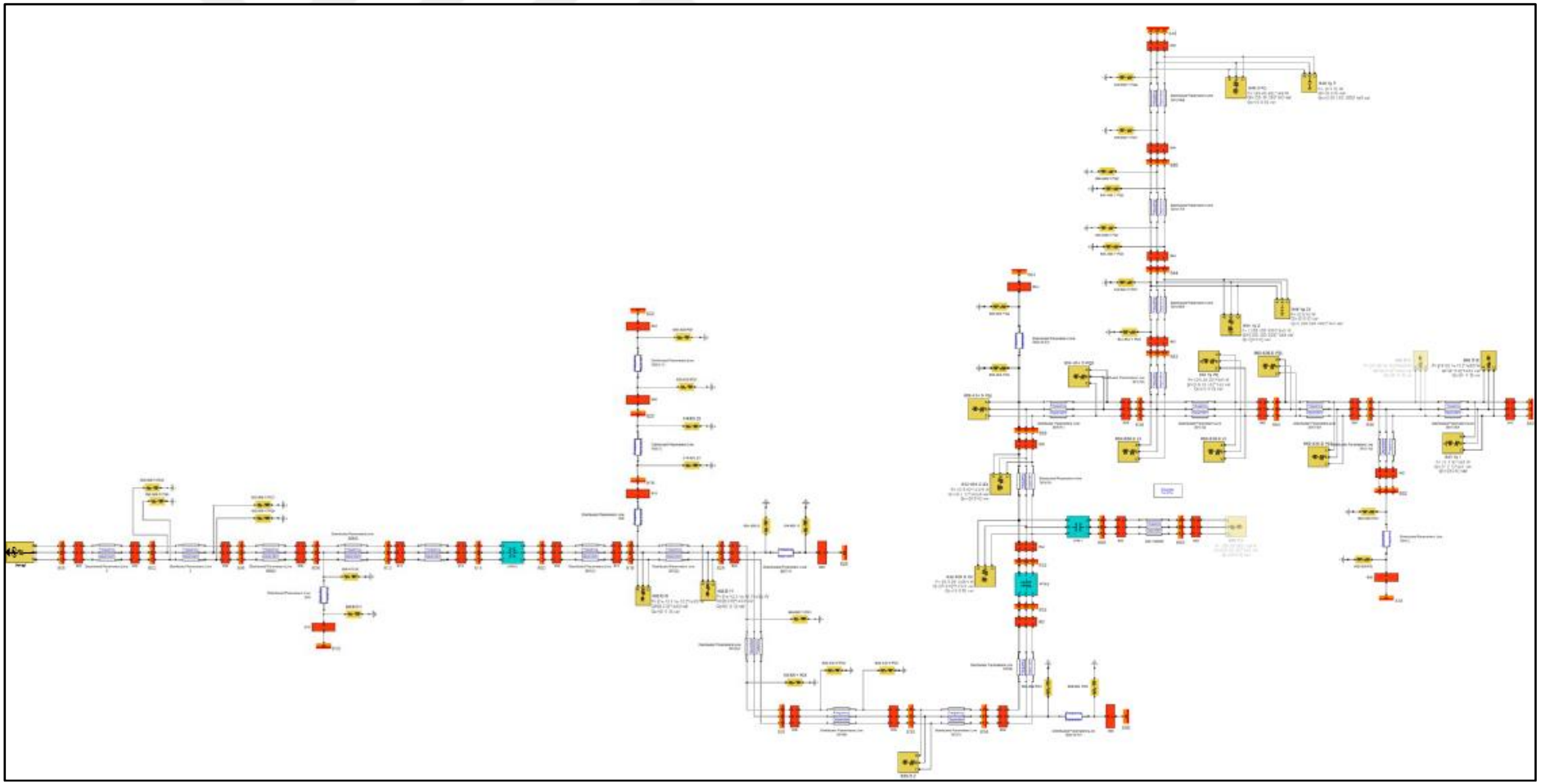


Figure A.1 : Modeled IEEE 34 bus test system in Simulink.

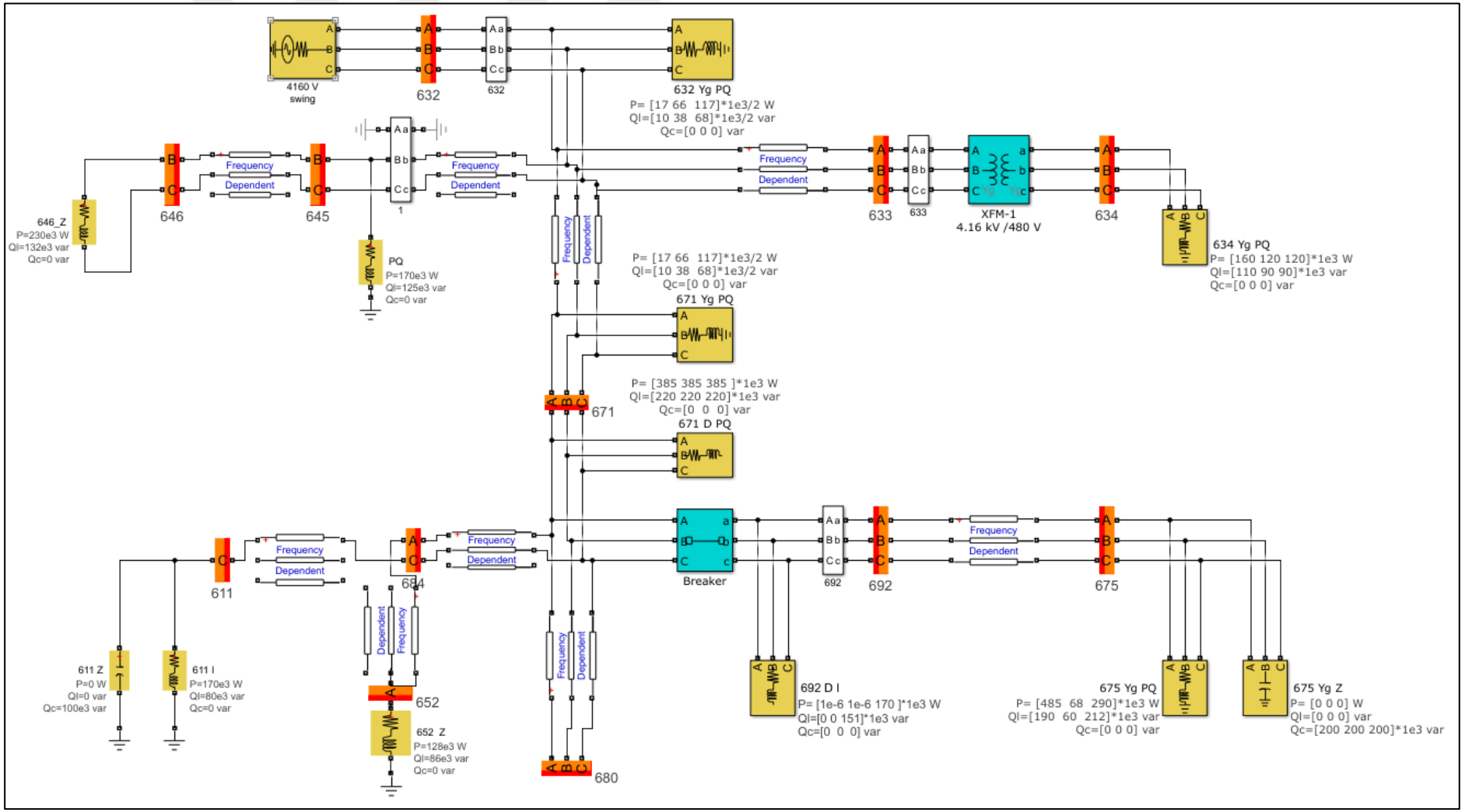


Figure A.2 : Modeled IEEE 13 bus test system in Simulink.

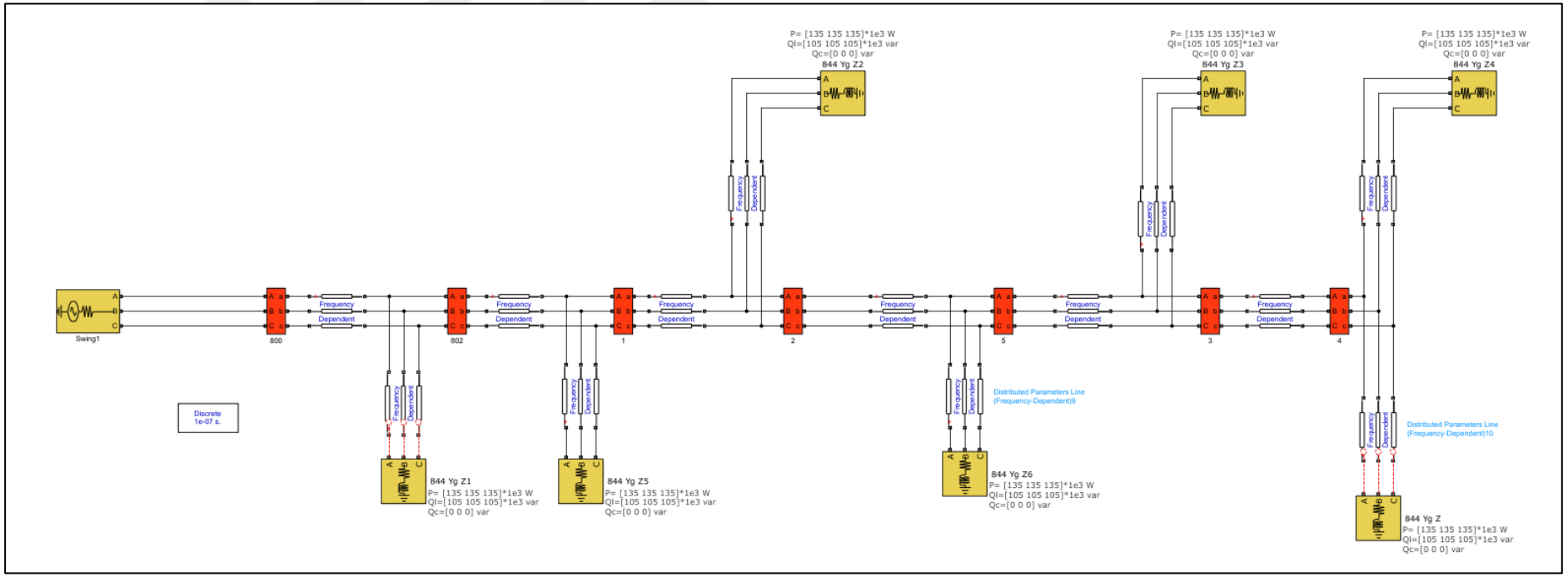


Figure A.3 : Modeled 14 bus test system in Simulink.

APPENDIX B: Verification of Modeled Networks

Frequency-dependent line models were employed in the simulations to achieve more accurate modeling. This detailed modeling incorporated phasing, spacing, and conductor information, including impedances, outside diameter, and geometric mean radius. An example of the line data from the IEEE 13 bus test system is presented in Figure B.4. Based on the entered parameters, the resistance (R), inductance (L), and capacitance (C) values were calculated per kilometer at 60 Hz. These computed parameters were compared with the published ones listed in Table B.1 to verify the accuracy of the impedance values. This comparison reveals that the two sets of values are very similar, indicating that the modeling of the line parameters is correct. Although one example is provided in this section, the same verification was conducted for all line models.

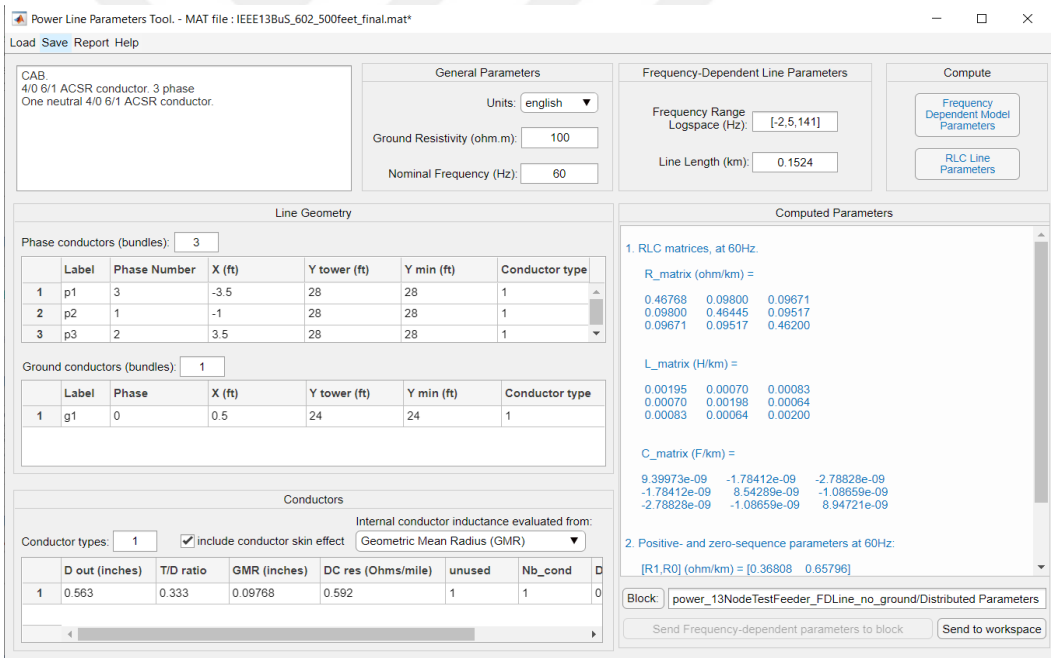


Figure B.4 : An example of line parameters for frequency dependent line models.

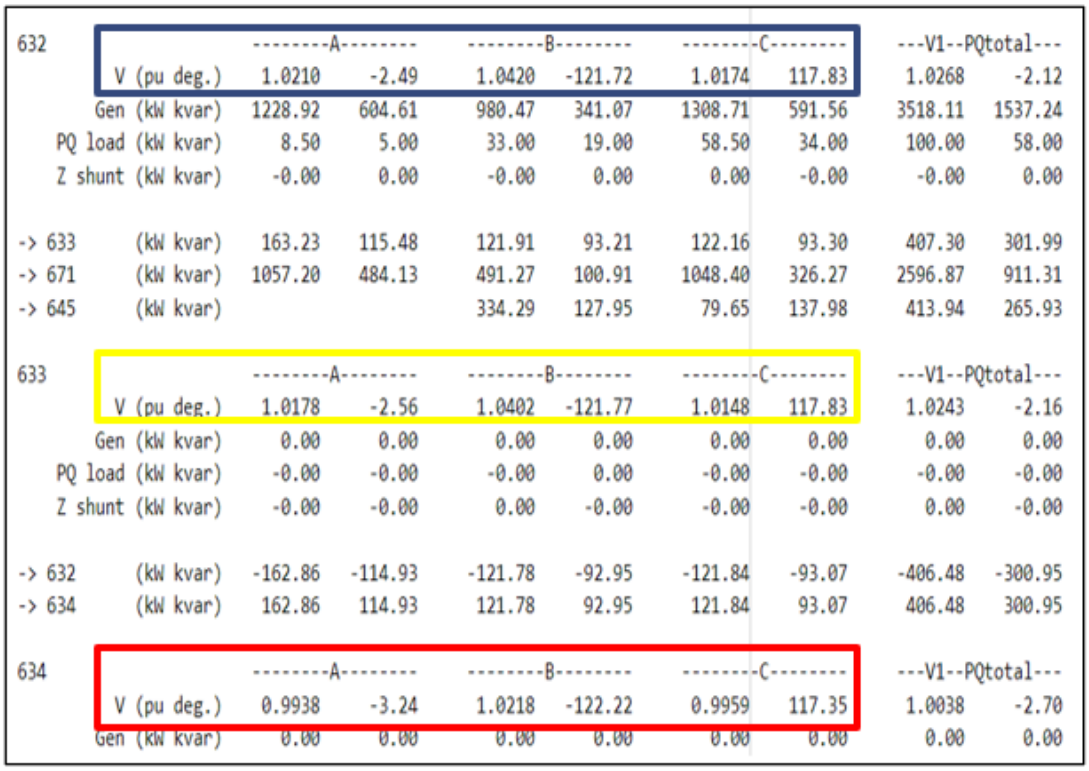
Table B.1 : Published RLC values from the literature.

R Matrix (ohm/km)	0.4676	0.0982	0.0969
	0.0982	0.4645	0.0954
	0.0969	0.0954	0.4621
L Matrix (L/km)	0.0019	0.0007	0.0008
	0.0007	0.0020	0.0006
	0.0008	0.0006	0.0020
C Matrix (F/km)	9.393e-9	-1.783e-9	-2.786e-9
	-1.783e-9	8.537e-9	-1.086e-9
	-2.786e-9	-1.086e-9	8.941e-9

Upon validating the impedance, the confirmation of the network modeling was conducted by comparing the load flow results for all the studied distribution networks. Figure B.5 and Figure B.6 provide an example comparison of the IEEE 13-bus test system. As depicted in the figures, the voltage measurements at the same buses, calculated using the model with frequency-dependent lines and the pre-existing model utilizing pi lines, are nearly identical. This finding verifies the precision of the modeling.

		--------A--------		--------B--------		--------C--------		---V1--PQtotal---	
632	V (pu deg.)	1.0210	-2.49	1.0420	-121.72	1.0174	117.83	1.0268	-2.12
	Gen (kW kvar)	1228.76	604.57	980.42	340.88	1308.75	591.42	3517.93	1536.86
	PQ load (kW kvar)	8.50	5.00	33.00	19.00	58.50	34.00	100.00	58.00
	Z shunt (kW kvar)	-0.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00
-> 633	(kW kvar)	163.22	115.45	121.93	93.21	122.14	93.32	407.29	301.98
-> 671	(kW kvar)	1057.04	484.12	491.20	100.72	1048.45	326.11	2596.69	910.95
-> 645	(kW kvar)			334.29	127.94	79.66	137.99	413.95	265.93
633		--------A--------		--------B--------		--------C--------		---V1--PQtotal---	
	V (pu deg.)	1.0180	-2.55	1.0401	-121.77	1.0148	117.83	1.0243	-2.16
	Gen (kW kvar)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	PQ load (kW kvar)	-0.00	-0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00
	Z shunt (kW kvar)	-0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00
-> 632	(kW kvar)	-162.86	-114.93	-121.78	-92.95	-121.84	-93.07	-406.48	-300.95
-> 634	(kW kvar)	162.86	114.93	121.78	92.95	121.84	93.07	406.48	300.95
634		--------A--------		--------B--------		--------C--------		---V1--PQtotal---	
	V (pu deg.)	0.9939	-3.23	1.0217	-122.22	0.9959	117.35	1.0038	-2.70
	Gen (kW kvar)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure B.5 : IEEE-13 buses network (pi line).



632		--------A--------		--------B--------		--------C--------		---V1--PQtotal---	
	V (pu deg.)	1.0210	-2.49	1.0420	-121.72	1.0174	117.83	1.0268	-2.12
	Gen (kW kvar)	1228.92	604.61	980.47	341.07	1308.71	591.56	3518.11	1537.24
	PQ load (kW kvar)	8.50	5.00	33.00	19.00	58.50	34.00	100.00	58.00
	Z shunt (kW kvar)	-0.00	0.00	-0.00	0.00	0.00	-0.00	-0.00	0.00
-> 633	(kW kvar)	163.23	115.48	121.91	93.21	122.16	93.30	407.30	301.99
-> 671	(kW kvar)	1057.20	484.13	491.27	100.91	1048.40	326.27	2596.87	911.31
-> 645	(kW kvar)			334.29	127.95	79.65	137.98	413.94	265.93

633		--------A--------		--------B--------		--------C--------		---V1--PQtotal---	
	V (pu deg.)	1.0178	-2.56	1.0402	-121.77	1.0148	117.83	1.0243	-2.16
	Gen (kW kvar)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	PQ load (kW kvar)	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00
	Z shunt (kW kvar)	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00
-> 632	(kW kvar)	-162.86	-114.93	-121.78	-92.95	-121.84	-93.07	-406.48	-300.95
-> 634	(kW kvar)	162.86	114.93	121.78	92.95	121.84	93.07	406.48	300.95

634		--------A--------		--------B--------		--------C--------		---V1--PQtotal---	
	V (pu deg.)	0.9938	-3.24	1.0218	-122.22	0.9959	117.35	1.0038	-2.70
	Gen (kW kvar)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure B.6 : IEEE-13 buses network (frequency dependent line).

CURRICULUM VITAE

Name Surname: Eren BAHARÖZÜ

EDUCATION:

- **B.Sc.:** 2015, Bahcesehir University, Faculty of Engineering and Natural Sciences, Department of Energy Systems Engineering
- **B.Sc.:** 2015, Bahcesehir University, Faculty of Engineering and Natural Sciences, Department of Electrical and Electronic Engineering
- **M.Sc.:** 2018, Bahcesehir University, Graduate School of Natural and Applied Sciences, Department of Electrical and Electronic Engineering

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Baharozu,** E., Ilhan, S., and Soykan, G. (2024). A Data-Driven-Based High Impedance Fault Location Method Considering Traveling Waves in Branched Distribution Networks, *IEEE Access*, *12*, 186535-186546.
- **Baharozu,** E., Ilhan, S., and Soykan, G. (2023). High impedance fault localization: A comprehensive review, *Electric Power Systems Research, 214, Part A.*
- **Baharozu,** E., Ilhan, S., and Soykan, G. (2023). Evaluation of the Effects of Noise and Sampling Rate on Detection of High Impedance Fault with Machine Learning Methods on the Distribution System, *2023 IEEE PES GTD International Conference and Exposition (GTD)*, Istanbul, Turkiye, May 22-25.